


```
Score: 190.50 Matches: 52
Percent Similarity: 45.76% Conservative: 29
Best Local Similarity: 29.38% Mismatches: 53
Query Match: 8.47% Indels: 43
DB: 29 Gaps: 7

US-09-985-689A-6 (1-434) x TA315H10P (1-574)
QY 18 GlyLeuTyrGlyGlnGlyGlnValValAlaValAlaAlaAspThrGlyLeuAspThrGlyArg 37
DB 63 GGTATTGCGGTAGTGGTGAATAGGGTATCCGATACGGGTATCGCATTT-----116
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaIleTyrAlaLeuGly 57
DB 117 AACAGCTGTTCTCCACGATCCA-----AATCAAGAGGTGGCGCTTATCCCGAAGTT 170
QY 58 ArgThrAsnAsn-----61
DB 171 AACTATAACCCCGCAAAATCGTGTCACTTCCCGCGTGTGACTTCATCCCGGGGATAC 230
QY 62 -----AlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
DB 231 TTTCGTGGGGATGAGGAATAGGTATGCGACGACGACGTGGCAGTACCGCAGCGGGAGT 290
QY 80 GlyThrSerAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93
DB 291 GTTATTAGTAACGACGGTAATGCCAAGTATAATGTTGTCGCAAGGGGGCGAAGATTTC 350
QY 94 PheGlnSerValMetAspSerAsnGlyGlyLeuGly-----Gly 106
DB 351 TTCAGG-----GGTTGGCTGCCATCCAGTCAGACCTGTC 389
QY 107 LeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHis 126
DB 390 CTCGCCACGACGTTACTCAATATTCTGTCGCCGATATGGCTGGAGCGCGTGTGTC 449
QY 127 ThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAspSerArgAsnValAsp 146
DB 450 TCAAACTCGTGGGTTTGTCTCCCTCCCGAGTATCTGCTGGGAAAGATATGGAT 509
QY 147 AspTyrValArgLysAsnAspMetAlaValLeu-----PheAlaAlaGlyAsn 162
DB 510 GAGTTTGGAGTAGTATGACGATGCGGTACTTATCTTCTCCACTGGCAAC 560

RESULT 12
BJ369190 633 bp mRNA linear EST 08-MAR-2002
LOCUS BJ369190 Dictyostelium discoideum cDNA library, CF Dictyostelium
DEFINITION discoideum cDNA clone ddc49116 5', mRNA sequence.
ACCESSION BJ369190
VERSION BJ369190.1 GI:19278573
KEYWORDS Dictyostelium discoideum
SOURCE Dictyostelium discoideum
ORGANISM Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 633)
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination
stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers
1. .633
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"

FEATURES
source
```

```
ORIGIN
Alignment Scores:
Pred. No.: 7,22e-09 Length: 633
Score: 189.00 Matches: 55
Percent Similarity: 41.55% Conservative: 31
Best Local Similarity: 26.57% Mismatches: 49
Query Match: 8.40% Indels: 72
DB: 12 Gaps: 6

US-09-985-689A-6 (1-434) x BJ369190 (1-633)
QY 136 GlyAlaTyrThrAspSerArgAsnValAspTyrValArgLys---AsnAspMet 154
DB 6 GGTGGTTATTCGATGATGCTGGTGTATGTATGATTCCTCTGATGATCCAGAAATTC 65
QY 155 AlaValLeuPheAlaAlaGlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGly 174
DB 66 TCTATACTAAGAGCTGCTGTAAT---AACGAGCTATTTCATCTTTATTAGCTCAAGCA 122
QY 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
DB 123 ACAGCTAAAAATGCAATTACAGTTGGTGTGCTGAGCAACAGCTCATGTAAATTATGTGCA 182
QY 195 -----TyrAlaAspAsnIle-----199
DB 183 GATGCATTGGAATATTATGATTTCTCAGATAATGCTAATTTTCAAAGACCATTGTTATTC 242
QY 199 -----199
DB 243 GATAAGAGATTGTTAATTATACGACCGCTAAATGTTGCTCAGAGTTTCAAATGTTAAA 302
QY 199 -----199
DB 303 GGTTTACAATTATGTTGTCAGCATCTATTAAACAAATGTCATCGGATTCATTCAACA 362
QY 200 -----AsnHisValAlaGlnPheSerSerArgGlyProThrLys 212
DB 363 CAACCTCAATTTTATTAATGAAATATATGCGGATCATTTCTCAAAAGGTCACACAT 422
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
DB 423 GATGGTAGATTGAACCTGATATAGTTGCACTGCTGATATATATACATCGCAAGATCA 482
QY 233 Ser-----LeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 246
DB 483 AATGCTGAGAAATTCACAGACCAATGCTGATGTTCTTTA-----CCAAATGCAAT 536
QY 247 LysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 266
DB 537 GGTCTAATGCAATATCTGTCATCATGCAATGCAACACCATTCGTCACAGCAACAACA 596
QY 267 GlnLeuArgGluHisPheIle 273
DB 597 ATCTTTAGACAATATTAGTT 617

RESULT 13
BJ750157 718 bp mRNA linear EST 25-SEP-2001
LOCUS BJ750157 FG02_10G08_R FG02_AAF_C_ECORC Fusarium graminearum_mycelium
DEFINITION Gibberella zeae cDNA clone FG02_10G08, mRNA sequence.
ACCESSION BJ750157
VERSION BJ750157.1 GI:15771959
KEYWORDS EST.
SOURCE Gibberella zeae
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 718)
```

AUTHORS Harris,L.J., Glasco,T., Rocheleau,H., Allard,S., Chapados,J., Couroux,P., De Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A., Spcott,D. and Rinker,N.A.
TITLE Expressed Sequence Tags from Fusarium graminearum mycelium
JOURNAL Unpublished (2001)
COMMENT Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
Bldg 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris@em.agr.ca.
FEATURES Location/Qualifiers
1..718
/organism="Gibberella zeae"
/mol_type="RNA"
/strain="DAOM 180378"
/db_xref="taxon:5518"
/clone="Fg02.10G08"
/tissue_type="Mycelial tissue"
/dev_stage="Asexual"
/lab_host="E. coli (Sure cells)"
/clone_lib="Fg02 AAFPC ECORC Fusarium graminearum mycelium"
/notes="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
Site 2: XhoI; Mycelial tissue was collected from v8 agar
plates after a growth period of 6-7 days at 25 C with 14
hrs (FL/UV) day light exposure. Mycelia was ground in
liquid nitrogen prior it's storage at -80 C until RNA
extraction. Directional cloning with 5' end of cDNA cloned
into EcoRI site of pBluescript and 3' end of cDNA cloned
into XhoI site of pBluescript (Stratagene, La Jolla, CA)."
ORIGIN
Alignment Scores:
Pred. No.: 8.98e-09 Length: 718
Score: 189.00 Matches: 66
Percent Similarity: 43.05% Conservative: 30
Best Local Similarity: 29.60% Mismatches: 89
Query Match: 8.40% Indels: 38
DB: 12 Gaps: 10
US-09-985-689A-6 (1-434) x B1750157 (1-718)
QY 46 PheArgGlyLysIleThrAlaIleTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspPro 65
Db 9 TTCAGAGSTGCTGCTCAGGCTGTCTACATCGCTTCAGCGGCCAGACGCT---GACACC 65
QY 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-----GlyThrSerAsn 83
Db 66 AACGGTCACGGAACCTACGTTGCTGGCATTGTCGGGAAGACATACGCGTGTGCCAAG 125
QY 84 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGly--- 102
Db 126 AAGGCCACCATC---CAAGCTGTCAAGTCTTCAGGGTAGTTCATCCAGCCTCCATC 182
QY 103 GlyLeuGlyGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTyrSerAlaGly 122
Db 183 ATCTCTCGCTTCACTGGCTGCGTCAACAGCATCATCTCCAAG----- 227
QY 123 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSer 142
Db 228 GGCAGAACCAAGACATCAGTCAGTCGTCATATGCTCTCGCGGTGGTACTCTGCTCTCTC 287
QY 143 ArgAsnValAspAspTyrValArgLysAsnAspMetAlaValLeuPheAlaAlaGlyAsn 162
Db 288 AACAAAGCTGTCGAGTCKYTCAGCTCCGGTATTATCTCTGCCATTGTCGCGGTAC 347
QY 163 GluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrVal 182
Db 348 GATGGTGCAACGCTGCGCAACACTTCT---CCTGCTCTCTCTCCAGCGCCAWCACTGTC 404
QY 183 GlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisVal 202

Db 405 GGTGCCATTGACAGC-----AATGGGCC-----ATT 431
QY 203 AlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspValMetAla 222
Db 432 KCCTCGTACTCCAACTACGTCACGTTCTC-----GATACTTTGCT 473
QY 223 ProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAla 242
Db 474 CCTGCACCAAGCGTTCTCTCCGCC-----TGCTAC 503
QY 243 AsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAla 262
Db 504 ACCAGCAACAGTCCACCAACACCATCAGCGGACACCTCCATGCTACTCCACATTCGCC 563
QY 263 GlyAsnVal 265
Db 564 GGAATTGTC 572
RESULT 14
BZ893395/c
LOCUS
DEFINITION HL2 0177 H1 pUC18 Library Halorubrum lacusprofundi genomic 5',
genomic survey sequence.
ACCESSION BZ893395
VERSION BZ893395.1 GI:33343985
KEYWORDS GSS.
SOURCE Halorubrum lacusprofundi
ORGANISM Halorubrum lacusprofundi
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.
REFERENCE 1 (bases 1 to 716)
AUTHORS Goo,Y., Roach,J., Glusman,G., Baliga,N.S., Deutsch,K., Pan,M.,
DasSarma,S., Ng,W.V. and Hood,L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
FEATURES Location/Qualifiers
1..716
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:2247"
/clone_lib="H1 pUC18 Library"
/note="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Halorubrum lacusprofundi genomic DNA
using pUC18/SmaI/BAP plasmid"
ORIGIN
Alignment Scores:
Pred. No.: 1.82e-08 Length: 716
Score: 186.00 Matches: 72
Percent Similarity: 48.18% Conservative: 34
Best Local Similarity: 32.73% Mismatches: 77
Query Match: 8.27% Indels: 38
DB: 28 Gaps: 10
US-09-985-689A-6 (1-434) x BZ893395 (1-716)
QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 701 TCCGATCGCTCGAGCAGNAGCGGACCGCACTACGTCGCGGACCGTCCGCGCGGA 642
QY 80 GlyThrSerAsnLys-----GlyMetAlaProGlnAlaAsnLeuValPheGlnSerVal 97
Db 641 AACGGAGTGGGACCGCAATCGCTCGACCGAAGCGGTGCTCCACGGATCAAGGTG 582

```
Qy 98 MetAsp-SerAsnGlyLeuGlyLeuProSerAsnValSerThreuPheSerG1 117
Db 581 TTCAGCAGCAGCGCA
Qy 117 nAlaTySerAlaGlyAlaArgileHisThrAsnSerTirP-
Db 536 GGAATGGAACAGCGGAGATCCGAGTCGACGTGCTTCCAGTACGACCTCGGTGGC 477
Qy 134 lAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyxValArgLysAsnAspMe 154
Db 476 GACGGGCATACCATCTTCCATCGAACCG-GTTCGGAACACTCGCAGTCCCGGAAGAT 418
Qy 154 tAlaValLeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThreLeuSerAlaProG1 174
Db 417 CGCGGTGCTT--TCGCGCGGGAACATCGGTCAA-----GGACGTCGAGTCTTCCCG 367
Qy 174 yThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg-----ProSerPh 192
Db 366 GAACGCTACGACTCGTCGGTTCGGAGCGGTCAACGACGCGCGCGTCCGCGCACTT 307
Qy 192 eGlySerTyThrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThr-- 211
Db 306 CTCGAGC---GGAGACGAGATCAACACGTCACGTCGAGCGGTGGGAAGTATGCCCGCGGGA 250
Qy 212 ---LysAspGlyArgileLysProAspValMetAlaProGlyThrPheIleLeuSerAl 230
Db 249 CTGCCCCGACGAGTACGCGTTCGAGCGTTCGCGCCAGCGCGTACGATCTCGGC 190
Qy 230 aArg-----SerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLy 247
Db 189 GGAACCGCGCGGTACAAACATCCGAAGGAC----- 159
Qy 247 styAlaTyMetGlyTyThrSerMetAlaThrProIleValAlaGlyAsnValAla 266
Db 158 -----GGACCTCCATGCGCGCCGACGTCGAGCGCGTCCGCGGCG 117

RESULT 15
BQ770462
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BQ770462
665 bp mRNA linear EST 26-JUL-2002
UI-M-F70-BYV-m-19-0-UI-r1 NIH_BMAP_F10 Mus musculus cDNA clone
IMAGE:5702970 5', mRNA sequence.
BQ770462
BQ770462.1 GI:21978936
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 665)
NIH-MGC http://mgs.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue procurement: Dr. Jim Lin, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5,
Location/Qualifiers
1..665
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5702970"
/tissue_type="whole brain"
```

```
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (TI phase resistant)"
/clone_lib="NIH_BMAP_F10"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CAGCCACGAC. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
Program coordinator."
```

ORIGIN

```
Alignment Scores:
Pred. No.: 2,04e-08 Length: 665
Score: 185.00 Matches: 70
Percent Similarity: 43.93% Conservative: 35
Best Local Similarity: 29.23% Mismatches: 94
Query Match: 8.22% Indels: 40
DB: 13 Gaps: 9
```

US-09-985-689A-6 (1-434) x BQ770462 (1-665)

```
Qy 63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlyThrSer 82
Db 14 GATGATGGCTAGGCGCATGGCACATTCGTTCGAGGT---GTGATTCGACGATGAGGAG 70
Qy 83 AsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerValMetAspSerAsnGly 102
Db 71 TGCCAGGAGTTTGTCCAGATCGACATCTTCAGGGTCTTTACCAACAAT--- 127
Qy 103 GlyLeuGlyLeuProSerAsnValSerThrLeuPheSerGlnAlaTySerAlaGly 122
Db 128 -----CAGGTGCTTTACACATCTTCGTTCTGGATGCTTCACTATATGCC 172
Qy 123 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSer 142
Db 173 ATCTTAAGAAGATGGACGCTTCAACCTTAGCATCGGTGGCCGACCTTCATGATCAT 232
Qy 143 ArgAsnValAspAspTy-----ValArgLysAsnAspMetAlaValLeuPheAlaAla 160
Db 233 CCGTTTGTTCACAAGGTGTGGAAATTACACAGCTAACATGTAAATTATGTTTCTCTATT 292
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 293 GGCATATGATGGACCTCTCTATGGCACTCTGAATACCTCTGTCATCAGATGATGATG 352
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyThrAlaAspAsnIleAsn 200
Db 353 GGAGTGGGTGGCATTCGAC-----TTTGAAGATAACATC--- 385
Qy 201 HisValAlaGlnPheSerSerArgGly-----ProThrLysAspGly 214
Db 386 -----GCTCGCTTTCTTCCAGGGAATGACTACTCTGGAATTACACGAGGCTATGTT 439
Qy 215 ArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerLeu 234
Db 440 CGTGTGAAGCTGACATTCGTC-----ACCTATGTCGTGGAGTGGCGGTTCGCGT 490
Qy 235 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyThrAlaTyMetGlyTyThr 254
Db 491 GTGAAGGGGGGTGTC-----CGTCACTCTTCAGGGAGC 523
Qy 255 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheIleLys 274
Db 524 AGTGTGCTTCNCACAGTGTGTCGTGGCGCGTTCACCTTGTAGTAAGACAGTACAGAAG 583
```

Qy 275 AsnArgGlyTleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 293
Db 584 CGGAGCTGGTG-----AATCCTGCCAGTGTGAAGCAGCGTTGTAGCGTCAGCC 634

Search completed: April 4, 2004, 11:55:12
Job time : 2297.84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 4, 2004, 08:10:21 ; Search time 327.248 Seconds
(without alignments)
4948.852 Million cell updates/sec

Title: US-09-985-689A-7

Perfect score: 2252
Sequence: 1 NDVARGIVKADVAQNNFGLY.....EVOAYNPVSPQTPSLAIHV 433

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2466186 seqs, 1870095128 residues

Total number of hits satisfying chosen parameters: 4932372

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_31032004_161809_4271/app.query.fasta_1.3498
-DB=Published Applications NA -QFT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIS=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0995689.ecgn_1_1.601.grunat_31032004_161809_4271
-NCPU=6 -ICPU=3 -NO.MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSFLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/ptodata1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata1/pubpna/US09_PUBCOMB.seq:
10: /cgn2_6/ptodata1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata1/pubpna/US10_NEW_PUB.seq:
17: /cgn2_6/ptodata1/pubpna/US60_NEW_PUB.seq:
18: /cgn2_6/ptodata1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	2125.5	94.4	1305	15	US-10-385-662-1	Sequence 1, Appli
2	432.5	19.2	1977	13	US-10-090-624-11	Sequence 11, Appl
3	396	17.6	1236	13	US-10-090-624-2	Sequence 2, Appli
4	396	17.6	1962	13	US-10-090-624-15	Sequence 15, Appl
5	355	15.8	3624	14	US-10-156-761-5701	Sequence 5701, Ap
6	355	15.8	9025608	14	US-10-156-761-3306	Sequence 1, Appli
7	332	14.7	3417	14	US-10-156-761-1	Sequence 3306, Ap
8	325	14.4	1329	9	US-09-974-300-1934	Sequence 1334, Ap
9	325	14.4	1329	9	US-10-156-761-1	Sequence 1, Appli
10	323	14.3	135638	14	US-10-314-657-1	Sequence 5, Appli
11	311.5	13.8	4765	13	US-10-090-624-5	Sequence 113, App
12	297	13.2	1560	15	US-10-084-846A-113	Sequence 1, Appli
13	297	13.2	59816	15	US-10-084-846A-2	Sequence 2, Appli
14	297	13.2	59816	15	US-10-084-846A-1	Sequence 29, Appl
15	279.5	12.4	3743	10	US-09-927-827-23	Sequence 33, Appl
16	276.5	12.3	3788	10	US-09-927-827-33	Sequence 1, Appli
17	271.5	12.1	1306	9	US-09-966-921A-1	Sequence 5, Appli
18	271.5	12.1	1330	9	US-09-966-921A-5	Sequence 5384, Ap
19	259.5	11.5	3303	14	US-10-156-761-5384	Sequence 17, Appl
20	258	11.5	2166	12	US-10-344-231-17	Sequence 17, Appl
21	258	11.5	2166	12	US-10-363-332A-17	Sequence 1335, Ap
22	254.5	11.3	1971	9	US-09-974-300-1935	Sequence 112429,
23	254.5	11.3	2192	12	US-10-424-599-112429	Sequence 1338, Ap
24	253.5	11.3	1485	9	US-09-974-300-1938	Sequence 34, Appl
25	252	11.2	3884	10	US-09-927-827-34	Sequence 1, Appli
26	245.5	10.9	840	14	US-10-209-812-1	Sequence 11, Appl
27	245.5	10.9	1140	8	US-08-322-678-11	Sequence 12, Appl
28	245.5	10.9	1140	8	US-08-322-678-12	Sequence 11, Appl
29	245.5	10.9	1140	16	US-10-323-324-11	Sequence 12, Appl
30	245.5	10.9	1140	16	US-10-323-324-12	Sequence 6, Appli
31	245.5	10.9	1143	14	US-10-313-853-6	Sequence 20, Appl
32	245.5	10.9	2588	12	US-10-344-231-20	Sequence 20, Appl
33	245.5	10.9	2588	12	US-10-363-332A-20	Sequence 30, Appl
34	236.5	10.5	3452	10	US-09-927-827-30	Sequence 18, Appl
35	232.5	10.3	2267	12	US-10-344-231-30	Sequence 18, Appl
36	232.5	10.3	2267	12	US-10-363-332A-18	Sequence 3, Appli
37	232	10.3	4338	9	US-09-891-711-3	Sequence 953, App
38	232	10.3	4338	12	US-10-342-887-953	Sequence 5689, Ap
39	230	10.2	1332	14	US-10-156-761-5689	Sequence 5, Appli
40	228	10.1	4198	9	US-09-891-711-5	Sequence 9, Appli
41	226	10.0	1140	15	US-10-146-905A-9	Sequence 6, Appli
42	224	9.9	1497	9	US-08-322-678-6	Sequence 1, Appli
43	224	9.9	1497	9	US-09-060-854B-1	Sequence 1, Appli
44	224	9.9	1497	14	US-10-033-325-1	Sequence 1, Appli
45	224	9.9	1497	14	US-10-228-572-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US2004002432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAITO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAEKI, KATSUHISA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASATUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10/385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

;; PRIOR FILING DATE: 2002-10-18
;; NUMBER OF SEQ ID NOS: 20
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 1
;; LENGTH: 1305
;; TYPE: DNA
;; ORGANISM: Bacillus sp. KSM-KP43
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1305)
;; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 8,02e-222 1305
Score: 2125.50
Percent Similarity: 97.93%
Best Local Similarity: 93.55%
Query Match: 94.38%
DB: 15
Gaps: 1
Indels: 1
Mismatch: 8
Conservative: 19
Matches: 406
Length: 1305

US-09-985-689A-7 (1-433) x US-10-385-662-1 (1-1305)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 1 AATGATGTGGCGTGGAAATGTGCAAGCGATGTGGCTCAGAGCAGCTACGGTTGTAT 60
Qy 21 GlyGlnGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGACAAGGACAGATCGTAGCGGTGGCCGATACAGGGCTTGATACAGTCGCAATGACAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCGATGATGAGGCTTCGCGGGAATATTACTGCAATTATGATGATGGGACGAGCAAT 180
Qy 61 AsnAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 181 AATGCCAATGATACGAATGGTTCATGGTACGATGGCTGGCTGGCTGGCTGGCTGGCTGGCT 240
Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 241 TCCACTAATAAGAAATGGCCCTCAGCGNAATCTAGTCTCCATCTCATGATGATAGC 300
Qy 100 GlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
Db 301 GGTGGGGACTTGGAGGACTACCTTCGAATTCGAAACCTTATTTCAGCCAAAGCATACAGT 360
Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139
Db 361 GTGTGTCCAGAAATTCATCAAACTCTCGGAGCAGCAGTGAATGGGCTTACACAACA 420
Qy 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
Db 421 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCTTTTCGTGCC 480
Qy 160 GlyAsnGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 481 GGGAAATGAAGACCGAACCGGCGAACCATGATGCGCAGCAGCAGCTAAATAATGCAATA 540
Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
Db 541 ACAGTCGGAGCTACGGAAACCTCGCCCAAGCTTTGGTCTTATGCGACAATATCAAC 600
Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
Db 601 CATGTGGCAGACTTCTTCAGTGGACCGCAAAAGGATGACCGATCAAAACCGGATGTC 660
Qy 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
Db 661 ATGGCACCGGGACGTTCTATCTATCATCATCATCATCTCTCTGACCGGATTCCTCTTC 720
Qy 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259

Db 721 TGGCGAACCATGACAGTAATAATGATCATATGATGGTGGACGTCCTGCTACACCATC 780
Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
Db 781 GTTCTGGAACAGTGGCAGCAGCTTCGTGAGCAATTTCTGAAAAACAGAGCATCACACCA 840
Qy 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db 841 AAGCCTTCTCTATTAAAGCGGCACTGATTTGCCGTCGACGCTGACATCGGCTTGGCTAC 900
Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db 901 CCGAACGGTAACCAAGGATGGGACGAGTGACATGGATAAATCCCTGAAACGTTGCTAT 960
Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339
Db 961 GTGAACGAGTCCAGTCTCTCTATCCACGAGCAAAAGCAGCTACTCTGTTTACTGCTACT 1020
Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359
Db 1021 GCCGCAAGCCTTTGAAAATCTCCCTGTTGTTCTGATGCCCTGGGAGCACAACTGCT 1080
Qy 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379
Db 1081 TCCGTAAACGCTTGCAATGATCTGGACCTTGCTATTACCGCTCCAAATGGCACACAGTAT 1140
Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db 1141 GTAGAAATGACTTACTTCCCATACATGATTAATGATGGATGGCGCAATAACGTAGAA 1200
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
Db 1201 AATGATTTATTAAATGACCAAAAGCGGACGATTAACAATTGAGGTACAGGCTTATAAC 1260
Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 1261 GTACCGGTGGACCAACAGACCTTCTCTGTTGCAATTTGTAAT 1302

RESULT 2

US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOUO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 6.62e-37
Score: 432.50
Percent Similarity: 44.85%
Best Local Similarity: 30.66%
Query Match: 19.21%
DB: 13
Length: 1977
Matches: 134
Conservative: 62
Mismatch: 156
Indels: 85
Gaps: 15

```
US-09-985-689A-7 (1-433) x US-10-090-624-11 (1-1977)
QY 8 ValLysAlaAspValAlaGlnAsnAsnProGlyLeuTyrGlyGlnGlyValAla 27
Db 433 ATAGGGCGGATACCGTCTGGAACTCCCTCGGCTACGACGAGCGGTGTGGTGGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATACAGCGAAC-----CACCCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GGCACAGGTATAGGCTGGTACGACCGCGCTCAACGCGAGGTGACCCCTACGATACGACAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82
Db 595 GGCACAGGAACCCACGTTCCGGGTATCGTTCCGGAACCGCGAGCGGTAACTCCCGATAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100
Db 655 ATAGGCGTCCGCGCGCGCGAGGCTCGTCGGCGTCAAGGTTCTCGGTGCGCAGCGTTCCG 714
QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
Db 715 GGAAGCGTCCACCATCATCGCGGTGTGACTGGTCTGCCAGAACAGGACAAAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp 140
Db 775 GGGATAAGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGCGAACCGACTCC 834
QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
Db 835 CTCAGTCAGCGCGTCAACCAACGCTGGAGCGCGGTATAGTAGTCTGGTCCGCCCGCGC 894
QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 895 AACAGCGGCGGAACACCTACCGTCTCCCGCGCGCGCGAGCAAGGTCTATAACC 954
QY 181 ValGlyAlaThrGlnAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
Db 955 GTCGGTGCAC-----GTTGACAGCAACGACCAAC 981
QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyValGlyLeuAspValMet 220
Db 982 ATCGCCAGCTTCTCCAGCGGGGACCGACCGCGGAGGCTCAAGCCGGAAGTCTGTC 1041
QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 1042 GCCTCCGCGGTTGACATCATAGCCCGCGCGCGCAGC-----GGAACCAAGCATGGGC 1092
QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCCGATTAACGACTACTACCAAGGCTCTGGNACCAAGCATGGCCACCCCGCACGTT 1152
QY 261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
Db 1153 TCGGGCGTTGGCGCTCTATCTCTCCAGGCGCCAC-----CCG 1188
QY 280 Lys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
Db 1189 AGCTGGACCCCGGACCAAGGTGAAGACCGCCCTCATCGACCGCGGCACATAGTCCGCCCC 1248
QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyLysGlnGlyTrpGlyArgVal 309
Db 1249 AAGGAGATACGGACATCGCTACGTTGCG-----GGTAGGTTG 1287
QY 310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326
Db 1288 AACGTCCTACAGCCCATAGTACGACGAGTACGCGCAAGCTCACCTTCACCGGCTCGGTC 1347
QY 327 SerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346
Db 1348 GCCGACAAAGGAAGCGCCACCCACACCTTC----- 1377
QY 347 SerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364
Db 1378 -----GACGTACGCGCGCCACCTTCTGGTACCCGCCCTCTACTGGGAC 1422
QY 365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381
Db 1423 ACGGCTCGAGCGACATCGACCTCTACTCTACGACCCCAACGGAACGAG---GTTGAC 1479
QY 382 AsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401
Db 1480 TACTCTTACACCGCTTACTAC-----GGCTTCGAGAAGGTC 1515
QY 402 PheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyr 418
Db 1516 GGCTACTACAAACCGACCGCGCGGAACCTGGACGCTCAAGGTCGTGAGTAC 1566

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIORITY FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2
Alignment Scores:
Pred. No.: 3,22e-33 Length: 1236
Score: 396.00 Matches: 127
Percent Similarity: 42.56% Conservative: 56
Best Local Similarity: 29.53% Mismatches: 163
Query Match: 17.58% Indels: 84
DB: 13 Gaps: 15

US-09-985-689A-7 (1-433) x US-10-090-624-2 (1-1236)
QY 16 AsnPheGlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTTGGGATATGATGTTCTTGGAAATCAATAAGGAATAATTCACACTGGAATTGAC--- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAGAATGA----- 144
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 204
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
Db 205 ACTCATGTAGTCTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGCAAGTACCAAGGA 264
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGly 102
```



```

Db 265 ATGGCTCCAGAGTAAGCTGGCGGAATTAAGTTCTAGTCCGATGTTCTGGAAGC 324
Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
Db 325 ATATCTACTATAATTAAGGAGTTAGTGGCGGTTGATAACAAGATAAGTACGGAATT 384
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg 142
Db 385 AAGGTCAATTAATCTTTCTTGGTTCAAGCAGCTCAGATGGTACTGACGCTCTAAGT 444
Qy 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTTAATGACGCGTGGATGCTGATTAAGTCTGCTGGCGTGGAAACAGT 504
Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTAACAGTATACAACTGGTTCTCCAGCAGCTCAAGCAAGTATTATACAGTTGA 564
Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202
Db 565 GCC-----GTTGCAAGATGATGATTATAACA 591
Qy 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGCCAACTGCAGCGGCGCTTAAGCGTGAAGTGTGTGCTCCA 651
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerPheTrpAlaAsn 242
Db 652 GGAACCTGGATTAATGTGTCAGAGCAAGT-----GGAACCTAGCAGTGGTCAACA 702
Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCACTCCTCAGTAGTGGT 762
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
Db 763 ATTGACGCCCTCTGCTCCAA-----GCACACCCGAGCTGGACT 801
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
Db 802 CCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 861
Qy 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312
Db 862 GCCGATATAGCTACGGTGCA-----GGTAGGGTTAATGCATAC 900
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGlnLysThrSerProLeuSerThrSer 329
Db 901 AAGGTATAAACTACGATAACTATGCAAGAGCTAGTGTTCACCTGATATGTTGCCAACAAA 960
Qy 330 GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349
Db 961 GGCAGCCAACTCACAGTTCGTTATAGCGAGCTTCGTTGCTGTAACCTGCCACATTATAC 1020
Qy 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
Db 1021 TGGGCAATGCAAT-----AGCGACCTTGATCTT 1050
Qy 370 VallIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAsp 389
Db 1051 TACCTCTACAGTCCCAATGGAAACAG---GTTGACTACTCTTACACCGCTACTAT--- 1104
Qy 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
Db 1105 -----GGATTCGAAAAGTTGGTTATTACAACCCCACTGATGGA 1143
Qy 410 ThrTyrThrValGluValGlnAlaTyrAsn 419
Db 1144 ACATGSAATTAAGGTTGAAGCTACAGC 1173
RESULT 4
US-10-090-624-15
```

```

; Sequence 15, Application US/10090624
; Publication No: US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; US-10-090-624-15

Alignment Scores:
Pred. No.: 6,27e-33 Length: 1962
Score: 396.00 Matches: 127
Percent Similarity: 42.56% Conservative: 56
Best Local Similarity: 29.53% Mismatches: 163
Query Match: 17.58% Indels: 84
DB: 15 Gaps: 15

US-09-985-689A-7 (1-433) x US-10-090-624-15 (1-1962)

Qy 16 AsnPheGlyLeuTyrGlyGlnGlyValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACTTGGATATGATGGTTCCTGGATCAATAGGAATAATTGACACTGGAATTGAC--- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
Db 511 -----GCTTCTCATCCAGATCTCCAAAGGAAAAGTA----- 540
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGTGGGTAGATTTGTCAATGGTAGGATTAATCCATAGATGACCATGGACATGA 600
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
Db 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAAGGA 660
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGTTCTAGTGGCGATGTTCTGGAAGC 720
Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
Db 721 ATATCTACTATAATTAAGGAGCTGAGTGGCGGCTTGATTAACAAGATAAGTACGGAATT 780
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg 142
Db 781 AAGGTCAATTAATCTTTCTTGGTTCAAGCAGCTCAGATGGTACTGACGCTCTAAGT 840
Qy 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
Db 841 CAGGCTGTTAATGACGCGTGGATGCTGGATTAAGTATTAGTTGTTGTTGCTGGCAACAGT 900
Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGACCTAACAGTATACAACTGGTTCTCCAGCAGCTCAAGCAAGTATTATACAGTTGA 960
Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202
```

```
Db 961 GCC-----|||...:|
| 961 GCC-----|||...:|
| 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
| 988 AGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCCTTAAGCTTAAAGCTTGGTCTCA 1047
| 223 GlyThrTyrlleLysSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
| 1048 GGAACCTGGATTAATCTGCCAGAGCAAGT-----GGAACCTAGCATGGTCAACCA 1098
| 243 HisAspSerLysTyrlaTyrlMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
| 1099 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCACTCTCAGTACTGGT 1158
| 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
| 1159 ATTGAGCCCTCTTGTCTCAA-----GCACACCCGAGCTGGACT 1197
| 281 ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 292
| 1198 CCAGACAAAGTAAACACGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 1257
| 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGinglyTrpGlyArgValThrLeuAsp 312
| 1258 GCCGATATAGCCTACGCTGCA-----GGTAGGTTAATGCATAC 1296
| 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLysSerThrSer 329
| 1297 AAGGCTATAACTACGATAACTATGCAAGCTAGTGTCTCCTGGATATGTTGCCAACAA 1356
| 330 GlnLysAlaThrTyrlSerPheThrAlaGlnAlaGlyLysProLysLysIleSerLeuVal 349
| 1357 GGCAGCAAACTCACAGTTCGTATTAGCGAGCTTCGTTAACTGCCACATTATAC 1416
| 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
| 1417 TGGACAAATGCCAAT-----AGGACCTTGAATCTT 1446
| 370 ValIleThrAlaProAsnGlyThrLysTyrlValGlyAsnAspPheThrAlaProTyrlAsp 389
| 1447 TACCTCTACGATCCCAATGGAACAG--GTTGACTACTCTTACACCGCTACTAT--- 1500
| 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
| 1501 -----GGATTGAAAGTTGTTATTACAAACCACTGATGA 1539
| 410 ThrTyrlThrValGluValGlnAlaTyrlAsn 419
| 1540 ACATGGACAAATTAAGTTGTAAAGCTACAGC 1569
```

RESULT 5

```
US-10-156-761-5701
; Sequence 5701, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
```

```
LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. NO.: 4,49e-28 Length: 3624
Score: 355.00 Matches: 126
Percent Similarity: 44.47% Conservative: 51
Best Local Similarity: 31.66% Mismatches: 151
Query Match: 15.76% Indels: 70
DB: 14 Gaps: 17

US-09-985-689A-7 (1-433) x US-10-156-761-5701 (1-3624)

Qy 8 ValLysAlaAspValAlaGlnAsnAsn-----Phe 17
| 556 GTGAGGCCCAATGCGCGAGAGCAACGCCAGATCGTACGCGCGCGGTGGAGGCC 615
| 18 GlyLeuTyrlGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
| 616 GGGCTCACGGCGCAGCGCGTCAACCGTCCGCTGCGACACCGCGCGTGCACAC-- 669
| 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrlAlaLeuGly 57
| 670 -----ACTACCCCGGACCTCGCGCGCGGTGTCGCCGAGCAAGAGCTTCATC 717
| 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
| 718 GACGGGAGGAGGTGCGCGACCAACGCGCACGGGACCCACGTCACCTCGACCGTCGGC 777
| 78 Gly-----AsnAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
| 778 GGCAGCGCGCGCGCTCCGACGCGAGCGGCGGTGCGCGCGGTGCGCGCTCGCC 837
| 93 PheGlnSerIleMet---AspSerGlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGln 111
| 838 GTCGCAAGTGTCTACGCGACCAACGCGCGGGA-----ACCGAGTCCCAG 882
| 112 ThrLeuPheSerGlnAlaTyrlSerAla-----GlyAlaArgIleHisThrAsnSer 128
| 883 ATCATCGCGCGCATGGAATGCGCGCGCGGAGCGTGGTCCAGAGTCTGTCGATGAGC 942
| 129 TrpGlyAlaProValAsnGlyAlaTyrlThrThrAsp-----SerArgAsnValAspAsp 146
| 943 CTCGGATCG--ACCGAGGCCAGCGAGCGGACCGACCCCATGCCCGCGCGCTCGACACC 999
| 147 TyrlValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySer 166
| 1000 CTCTCCGAGGAGACCGCGCGCTCTTCTGCTGCGCGCGGAAACACCGGTGCCCTCC 1059
| 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186
| 1060 ---TCGATCGCTCGCGCGCGCGCGCTGCTGCGCTGACCTGCGCGCGC----- 1107
| 187 LeuArgProSerPheGlySerTyrlAlaAspAsnIleAsnHisValAlaGlnPheSerSer 206
| 1108 -----GTGACTCATCGACCGCGCGCGCTACTTCCACGAC 1143
| 207 ArgGlyProThrArg---AspGlyArgIleLysProAspValMetAlaProGlyThrTyrl 225
| 1144 GCCGCGCGCGCGCACCGCGGACAAAGCCCTCAAGCCGAGCTCGCCGACCGCGCTGCAC 1203
| 226 IleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 245
| 1204 ATCCGCGCGCGCGCTCCCAGCTCGCGCGCGCGCGCGCGCTAC----- 1245
| 246 LysTyrlAlaTyrlMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 265
| 1246 ---TACACCTCCATGAGCGGTACGTGATGCGACGCCCGCCCATGTCGCGCGGTGCGCGCG 1302
```

```
QY 266 GlnLeuArgGluHisPheValysAsnArgGlyValThrProLysProSerLeuLeuLys 285
Db 1303 CTCCTCGCGAGCAGACCCCGACTGAGCCGGCGCGGCTCAGGACCGCTGATGTC 1362
QY 286 AlaAla-----LeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsn 303
Db 1363 ACGTCCGAGCAACTCGACGCTCCGTATATACAGTTGGGGCGGGTCCGGTCAAGTGTGCGG 1422
QY 304 GlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaPhe----- 319
Db 1423 GACGCCGTGCGCGCCCGCTGACCGACCGGGCAGCGCGCCACCTCCGGCTTCACCCCTGG 1482
QY 320 -----ValAsnGluThrSerProLeuSerThrSerGlnLysAla 332
Db 1483 CCCCATGACGCGCGATGACCCGTCACGACGAGCGTCACTTCACTCACTCCCGACAG 1542
QY 333 ThrTyrSerPheThr---AlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpSer 351
Db 1543 ACGGTCCAGTTGAGCCTCGCGGTGCGGGCGCGCCCGCGGTGTCGCCACCCCTCGCGGAC 1602
QY 352 AspAla-----ProGlySerThrThrAlaSerLeuThrLeuValAsnAsp 366
Db 1603 ACCGCATCACCCTGCGCGCCCGCCACCGCAGCGCGCCGCCACCCGTCGACCGCGCGAC 1656

RESULT 6
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 3,47e-23 Length: 9025608
Score: 355.00 Matches: 126
Percent Similarity: 44.47% Conservative: 51
Best Local Similarity: 31.66% Mismatches: 151
Query Match: 15.76% Indels: 70
DB: 14 Gaps: 17

US-09-985-689A-7 (1-433) x US-10-156-761-1 (1-9025608)

QY 8 ValLysAlaAspValAlaGlnAsnAsn-----Phe 17
Db 6918813 GTCGAGCGGACATGCGCGGAGACACCGCGAGATCGGTACCGGGCGCGTGGGACGCC 6918872
QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 6918873 GGGCTCACGGGCGACGGCGCTCACCGTCCGCGGTGCTGCAGACACCGGGGTGACACC----- 6918926
```

```
QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 6918927 -----ACTCACCCCGACCTCCGCGCGGGTGTCCCGAGCAAGACTTTCATC 6918974
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 6918975 GACGGGAGGAGGTGCGCGACCGCAACCGGCACCGGACCCACCGTCACCTCGACCGTGGC 6919034
QY 78 Gly-----AsnAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 6919035 GGCAGCGCGCGCTCCGACGCGCAGCGCGCGCTGCGCGCGGTGCGACCTCGCC 6919094
QY 93 PheGlnSerIleMet---AspSerGlyGlyLeuGlyGlyLeuProAlaAsnLeuGln 111
Db 6919095 CTCGCAAGTGTCTCAGCAGCAGCGCGCGGGA-----AGCGAGTCCGAG 6919139
QY 112 ThrLeuPheSerGlnAlaTyrSerAla-----GlyAlaArgIleHisThrAsnSer 128
Db 6919140 ATATCGCGGGCATGGAATGCGCGCGGACGTGCGTGCAGCATCTCTCGATGAGC 6919199
QY 129 TrpGlyAlaProValAsnGlyAlaTyrThrThrAsp-----SerArgAsnValAspAsp 146
Db 6919200 CTCGGATCG-----ACCGAGCGCAGCAGCGGACCCCATGCGCGAGCGCTCGACACC 6919256
QY 147 TyrValArgLysAsnAspMetThrIleLeuPheAlaIleAlaGlyAsnGluGlyProGlySer 166
Db 6919257 CTCCTCGAGGAGACCGCGCCCTCTCTGCTGCGCGGGGAAACACCGGTGCCCTCC 6919316
QY 167 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 186
Db 6919317 ---TCGATCGGCTCGCGCGCGCGCCGACTCCGCGTGCAGCTCGCGCGCC----- 6919364
QY 187 LeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSer 206
Db 6919365 -----GTCAGCTCATCCGACCGCGCGCGCTACTTCCACGAGC 6919400
QY 207 ArgGlyProThrArg---AspGlyArgIleLysProAspValMetAlaProGlyThrTyr 225
Db 6919401 GCGGCGCGCGCGCACGCGGACAAACCCCTCAAGCGCGCAGCTCGCGCGACCCGCGGTG 6919460
QY 226 IleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 245
Db 6919461 ATCCGCGCGCGCGCTCCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919502
QY 246 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 265
Db 6919503 ---TACACCTCCATGAGCGGTACGTGATGCGGACGCGCCCATGTCCGCGGGTCCGCGCG 6919559
QY 266 GlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSerLeuLys 285
Db 6919560 CTCCTCGCGGACGACGACCCCGACTGGACCGCGCGCGCGCTCAAGACCGCTGATGTC 6919619
QY 286 AlaAla-----LeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsn 303
Db 6919620 ACGTCCGAGCACTCGACGCTCCGTATATCAGTTGGGGCGGGTCCGGTCAAGTGTGCGG 6919679
QY 304 GlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaPhe----- 319
Db 6919680 GACCGCGTGGCGCGCGCGCTCACCGCGAGCGCGCGCGCGCTCGGCTTCCACCGCTG 6919739
QY 320 -----ValAsnGluThrSerProLeuSerThrSerGlnLysAla 332
Db 6919740 CCCCATGACGCGCGATGACCCGTCACGAGCGGTCACTTCACTTCCAACTCTCCGACAG 6919799
QY 333 ThrTyrSerPheThr---AlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpSer 351
Db 6919800 ACGGTCCAGTTGAGCCTCGCGGTGCGGGCGCGCGCGCGGTGTCGCCACCCCTCGCGGAC 6919859
QY 352 AspAla-----ProGlySerThrThrAlaSerLeuThrLeuValAsnAsp 366
Db 6919860 ACCGCATCACCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919913
```

```
RESULT 7
US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURES:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 1,33e-25 Length: 3417
Score: 332.00 Matches: 137
Percent Similarity: 44.39% Conservative: 53
Best Local Similarity: 32.01% Mismatches: 161
Query Match: 14.74% Indels: 78
DB: 20 Gaps: 20

US-09-985-689a-7 (1-433) x US-10-156-761-3306 (1-3417)
Qy 18 GlyLeuTyrglyGinglyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyAyg 37
Db 745 GGGTACGACGCGCAGGGCGTCAAGTCCGCTCTGACACCGGTGTCGAC----- 795
Qy 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGly 57
Db 796 -----GGACCCACCGCGACCTCAAGGACCGAGTGGCGCGAGTCCAAAGAACTTCTCC 846
Qy 58 ArgThrAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 847 GCGCGCGCGCGACGCGCGGACCACTTCGGTCAAGGACGCGCGCGCGCGCGCGCGCG 906
Qy 78 Gly-----AsnAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 907 GGCACCGCGCGCAAGTCCACGCAAGTACAGGGTGTGCGCGCGCGCGCGCGCGCGCTC 966
Qy 93 PheGlnSerIleMetAsp-----SerGlyGlyLeuGlyLeuProAlaAsnLeu 110
Db 967 AACGCGAAGGTCTCGACGACACCGGTTCGCGGACGACTCCGCGCATCTCGCGCGCGATG 1026
Qy 111 GlnThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1027 GAGTGGCGCGCGAGCAG-----GGCGCGCGCGCTCGTCACTCGAGCTCGGCG 1074
Qy 131 AlaProValAsnGlyAlaTyThrAspSerArgAsnValAspAspTyValArgLys 150
Db 1075 -----GGCGCGGACACCGCGCGAGTCAAGCGGTGGAAGCGGAGGTCAACAG 1122
Qy 151 -----AsnAspMetThrIleLeuPhe-----AlaAlaGlyAsnGlyPro---Gly 165
Db 1123 CTCTCCGAGGAGAGGCGATCTCTTCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCG 1182
Qy 166 SerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGlu 185
```

RESULT 8

```
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
```

;; PRIOR APPLICATION NUMBER: JP 2001-272697

;; PRIOR FILING DATE: 2001-08-02

;; NUMBER OF SEQ ID NOS: 15109

;; SEQ ID NO 1

;; LENGTH: 9025608

;; TYPE: DNA

;; ORGANISM: Streptomyces avermitilis

;; FEATURE:

;; NAME/KEY: misc_feature

;; LOCATION: (4187715)

;; OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Alignment Scores:

Pred. No.: 1,12e-20 Length: 9025608

Score: 332.00 Matches: 137

Percent Similarity: 44.39% Conservative: 53

Best Local Similarity: 32.01% Mismatches: 161

Query Match: 14.74% Indels: 78

DB: 14 Gaps: 20

US-09-985-689A-7 (1-433) x US-10-156-761-1 (1-9025608)

Qy 18 GlyLeuTyrGlyClnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArg 37
 Db 4132505 GGTACGACGGAGGAGGCGCTCAAGATCGCGCTTCCTGGACACCGGTGTCGAC----- 4132455
 Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 4132454 -----GCGACCCACCGGACCTCAAGACCAAGTGGCGGAGTCCCAAGAACTTCTCC 4132404
 Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 4132403 GCGCGCGCGACGCCGCCGACCACTTCGGTCAAGCAGCAGCAGTCCGCTCCATCGCGCG 4132344
 Qy 78 Gly-----AsnAlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 92
 Db 4132343 GGCACCGCGCGCCCAAGTCCCAACGCAAGTACAAAGGCTGTGCGCGCGCGGCGGACCATCTC 4132284
 Qy 93 PheGlnSerIleMetAsp-----SerGlyGlyGlyLeuGlyLeuProAlaAsnLeu 110
 Db 4132283 AACGGCAAGGTCTCTGACACACACCGCGCTCCGCGGAGCTCCGGCATCTCGCGGCGCATG 4132224
 Qy 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
 Db 4132223 GAGTGGCGCGCGCGGAGCAG-----GGCGCGGAGCTCGTCAACCTGAGCTCGGC 4132176
 Qy 131 AlaProValAsnGlyValTyrThrThrAspSerArgAsnValAspAspTyrValArgLys 150
 Db 4132175 -----GGCGCGACACCCCGGAGATCGACCGCTGGAGCGGAGGTCAACAG 4132128
 Qy 151 -----AsnAspMetThrIleLeuPhe-----AlaAlaGlyAsnGlyPro---Gly 165
 Db 4132127 CTCTCGGAGGAGGAGGAGGATCTCTTCGCGATCGCGCGCGGCAACGAGGCGGAGTTCGGC 4132068
 Qy 166 SerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyValAlaThrGlu 185
 Db 4132067 GAGCAGACCATCGCTCCCGGCGGCGGCGGCGGCGGCTACCGTCTCGGCGCC----- 4132014
 Qy 186 AsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSer 205
 Db 4132013 -----GTGAACGACGACGACGACGACGATCGCTCTCTCTCC 4131981
 Qy 206 SerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTyr 225
 Db 4131980 AGCGCGGCGCGCGCTGACGCGCGCATCAAGCCGCGAGTCAACGACCGCGGCGTGGAC 4131921
 Qy 226 IleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 245
 Db 4131920 ATCACCAGCGCGC-----GGCGCGCGGCGGCGGCTCATCGACGAGGAGTGGCGAG 4131870
 Qy 246 Lys-----TyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262

Db 4131869 AAGCGGACGGCTACCTACCATCTCGGTACGTGCGACCGCGCATGTGCGGGCG 4131810
 Qy 263 AsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyValThrProLysProSer 282
 Db 4131809 GCGCGCGGATCTCTCAAGCAGCAGCAC-----CCCAACTGGTCTG 4131771
 Qy 283 Leu-----LeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
 Db 4131770 TTCGCGGAGCTCAAGGGCGCGCTG---ACGGCTCCCGGAGGGCGGCAAGTACACGCCG 4131714
 Qy 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
 Db 4131713 ---TTCCAGCAGGCGCTCGGCGGTATCGCGTTCGACAAGCGCATCAAGCAGTCCGTGATC 4131657
 Qy 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAl 340
 Db 4131656 -----GCCAACCCGAACTCGGTG-----AGCTTCGCGATCCAGCAG 4131621
 Qy 340 aGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSe 360
 Db 4131620 TGGCGCGACACCGCAGCAAGCCGCTCACCAGCAGCTACCTACCGCAACCTCGGCGAG 4131561
 Qy 360 rLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThr----- 377
 Db 4131560 AGTGACGTC-ACGCTGAACCTCGGTGCGGCGGCGACCAACCCCAAGGCGCTCGCGGTCC 4131502
 Qy 378 -----LysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAs 393
 Db 4131501 GTCGGCGCTTCTTCAAGCTCGGCGCGCAGCAAGGTCAACGCTCCCGCGC-----GG 4131454
 Qy 393 pGlyArgAsnAsnValGlu-----AsnValPheIleAsnAlaProGlnSerGlyTh 410
 Db 4131453 CGCGAAGGCTCGGTGCTGCTACCGTCAACGAGCTGGCGGCGCACCGGCGCGC 4131394
 Qy 410 rTyrThrValGluValGlnAla 417
 Db 4131393 GTACTCGCGTACGTGACGCGC 4131372

RESULT 9
 US-09-974-300-1934
 ; Sequence 1934 Application US/09974300
 ; Patent No. US20020146721A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Berka, Randy M.
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 ; TITLE OF INVENTION: Expression
 ; FILE REFERENCE: 10085.500-US
 ; CURRENT APPLICATION NUMBER: US/09/974,300
 ; CURRENT FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: 09/680,598
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/279,526
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 8481
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1934
 ; LENGTH: 1329
 ; TYPE: DNA
 ; ORGANISM: Bacillus licheniformis
 ; US-09-974-300-1934

Alignment Scores:
 Pred. No.: 1,98e-25 Length: 1329
 Score: 325.00 Matches: 105
 Percent Similarity: 43.98% Conservative: 41
 Best Local Similarity: 31.63% Mismatches: 118
 Query Match: 14.43% Indels: 68
 DB: 9 Gaps: 13

US-09-985-689A-7 (1-433) x US-09-974-300-1934 (1-1329)

Qy 11 AspValAlaGlnAsnAsnPheGlyLeuTyrGlyGlnIleValAlaValAlaAasp 30

```

; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Synthesizing Acyltransferases Associated with Type I Polyketide
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
; US-10-314-657-1

Alignment Scores:
Pred. No.: 2,55e-22 Length: 135638
Score: 323.00 Matches: 133
Percent Similarity: 41.63% Conservative: 56
Best Local Similarity: 29.30% Mismatches: 137
Query Match: 14.34% Indels: 128
DB: 14 Gaps: 22

US-09-985-689A-7 (1-433) x US-10-314-657-1 (1-135638)
QY 6 GlyLeuValValAlaAspValAlaGlnAsnAsnPhe----- 17
DB 8067 GCGAGGTTGAAGCGCATCTGCGCGACTCCACCCGCCAGATCCGCGCGAGAAAGGTATGG 8126
QY 18 -----GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
DB 8127 GCGAGGCGGACACCGCGCAGAGCTGAGTGCCTGCGATGCTCGACAGCGCGCGACAC 8186
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIle-----ThrAlaLeu 53
DB 8187 -----GAACACCGCGACCTGCTGCGCGAGGTGTCCGACAGCGCGCAGC 8228
QY 54 TyrAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAla 73
DB 8229 TTCTGTCCTCGCGAGGACGACATGCCC-----GACTACACGCGCCACGCGACGCTCGCC 8285
QY 74 GlySerValLeuGlyAsnAlaThr-----AsnLysGlyMetAlaProGln 88
DB 8286 TCGACCATCGTCGCGCAGCGCGCCTCCGACGCGAGGAGCGGGGTGTGCTCCCTCCGCG 8345
QY 89 AlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGly----- 102
DB 8346 GCGCGCGCTGCTCGCGCAGGTGCTCACTCCGAGGCGCGCGCGCGAGGATCTGGATC 8405
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
DB 8406 ATCGCGCGCATGGAG-----TGGCGCGCGCGCGCGAGCAGAGGCC 8444
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp----- 140
DB 8445 AGGATCATCAGCATGAGCTGCGC-----GCGCGCGGTGTCAGAGAACGCCCGGATG 8495
QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160
DB 8496 AGCCAGCGCTGCGAGAACTCAGCAGCAGCGCGCGGTGTTGTTGTTGATCGCGCGGCG 8555
QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB 8556 AACGCGCGCGCGCACTCC-----ATCAGCAGCGCGGTGCGCGAGACTCGCGCTGACC 8609
QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
DB 8610 GTGCGCGCC-----GTGCACTCCACCGACAGC 8636

```

RESULT 10
 US-10-314-657-1
 ; Sequence 1, Application US/10314657
 ; Publication No. US2003017588A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SHEN, Ben

201 ValAlaGlnPheSerArgGlyPro---ThrArgAspGlyValGlyLeuValProAspVal 219
8637 CTCGCCGACCTCTCCAGCGAGGCGCGCGGTGACGCGCGCGGTGAGACCGGAGATC 8696
220 MetAlaProGlyThrTyriLeuSerAlaArgSerSerLeuAlaProAspSerPhe 239
8697 ACCGCGCGCGCGGTGCGCATGCTCGCGCGCGGTGCGCATACAGCGCGCGGTGCGCAT 8756
240 TrpAlaAsnHisAspSerLysTyriAlaTyriMetGlyGlyThrSerMetAlaThrProIle 259
8757 TACACCAGTATGAGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGTGCGCAT 8795
260 ValAlaGlyAsnValAla-----GlnLeu 267
8796 GTCCGCGGTGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCG 8855
268 ArgGlnHisPheValLysAsnArgGlyValThrProLysProSerLeuLeuLysAlaAla 287
8856 AAGGAGGACACTGGTGCAGCGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGT 8912
288 LeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGly 307
8913 TACACCAGTATGAGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGTGCGCAT 8921
308 ArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerProLeuSer 327
8922 CGGCTCGAGCGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCG 8963
328 ThrSerGlnLysAlaAlaThrTyriSerPheThrAlaGlnAlaGlyLysProLeuLysIleSer 347
8964 ACCAGC-----ACCGCTACTCCGCG-----TTC 8987
348 LeuValTrpSerAspAlaProGlySerThr-----ThrAlaSerLeuThrLeuVal 364
8988 CACAGTGGCGCGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCG 9047
365 AsnAspLeuAspLeuValIleThr---AlaProAsnGlyThrLysTyriValGlyAsnAsp 383
9048 GCGGAGCGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGT 9104
384 PheThrAlaProTyriAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIle 403
9105 TTCAGCGCTCCGAGGACCAT-----GTC 9128
404 AsnAlaProGlnSerGlyThrTyriThrValGluValGlnAla 417
9129 ACCGTGCGCGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCGCGGTGCGCATGCGCG 9170

RESULT 11

US-10-090-624-5
Sequence 5, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Kunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090, 624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
PRIOR FILING DATE: 1997-06-10
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent in version 3.0
SEQ ID NO 5
LENGTH: 4765
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Synthetic
US-10-090-624-5
Alignment Scores:
Pred. No.: 3,7e-23 Length: 4765
Score: 311.50 Matches: 144
Percent Similarity: 38.00% Conservative: 57
Best Local Similarity: 27.22% Mismatches: 168
Query Match: 13.83% Indels: 161
DB: 13 Gaps: 20
US-09-985-689A-7 (1-433) x US-10-090-624-5 (1-4765)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 1024 GCGAATGGTTATGACATTCGATATGCGATATGCGATATGCGATATGCGATATGCGAT 1083
QY 41 -----SerMetHisGluAlaPheArgGly 48
DB 1084 GTTCACCTTGGCGAGTACACGTTACTTATGATGTTGCTGTTTGTAGCTACTACTACG 1143
QY 49 LysIleThrAlaLeuTyriAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn----- 66
DB 1144 CCTCTCAAC-----TACGTGCTTCGAGAAATA-----GATCCTAACCGAGAA 1185
QY 67 -----GlyHisGlyThrHisValAlaGlySerValLeu 77
DB 1186 TATGAGTATTTGGTGGGATGCGTACCGTACCGAATCAGTACGTTAGTGGAACTGTGCT 1245
QY 78 GlyAsnAlaThrAsn----- 82
DB 1246 GGTACGACGACCAACAATGATGCTTGGGATGCGTACGATGCTACTCTGTTGTAATGG 1305
QY 83 -----LysGly 84
DB 1306 GTGTTCTCAAGACTCTATGTTGGGATATACAGACGTTACACAGACACCGTGCAGGT 1365
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyLeuGly 104
DB 1366 GTTGCTCCAGGTGCCAATAATAGCAATACAGTCTTAGGAGTGATGA-----CG 1419
QY 105 GlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyriSerAlaGlyAlaArgIle 124
DB 1420 GTGAGCATGCGGATATATAGAGGTATGATGACGACGACCAACCGTGCAGACGTT 1479
QY 125 HisThrAsnSerTrpGly-----AlaProValAsnGlyValAlaTyriThrThrAspSerArg 142
DB 1480 ATAAGCATGATGCTCGTGGAAATGCTCCATCTTAGATGGTACTGATCCAGAAACGTT 1539
QY 143 AsnValAspAspTyriValArgLysAsnAspMetThrIleLeuPheAlaAlaGlnGlu 162
DB 1540 GCTGTGATGAGCTTACCGAAAGTACGCGTGTGTTGTTATTCGTATATAGCTGCGAAATGA 1599
QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
DB 1600 GGTCTCGCATTAACATCGTGTGGAAGTCTGTTGTTGCAACAAAGCAATACTGTGGA 1659
QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyriAlaAsp----- 196
DB 1660 GCTGCTGCA---GTGCCCATTAACGTTGGAGTTTATGTTTCCCAAGCACTTGGATATC 1716
QY 197 -----AsnIleAsnHisValAlaGlnPhe 204
DB 1717 GATTACTATGATTCATTACTTCCCGCGCTACACAAACGTT---AGATAGCATCTTCT 1773
QY 205 SerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThr 224
DB 1774 TCAGCAGAGCGCGAGATAGATGTTGAAATAAAACCAATAGTGTGCTCCAGGTAC 1833
QY 225 TyriLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAsp 244
DB 1834 GGAATTTTACTCATCCCTGCCGATGCGGAGTGGCGGAGCTGACTTC----- 1878

QY 245 SerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnVal 264
DB 1879 -----ATGTCGAACTTCGATGGCTACTCCACATGTCAGCGGTGCGTT 1923
QY 265 AlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSerLeuLeu 284
DB 1924 GCATCTCTCATAAGCGG---GCAAGGCGGAGGGAATATACTACATCCAGATTAAT 1980
QY 285 LysAlaAlaLeuIleAlaGlyAlaAla-----AspValGlyLeuGly 298
DB 1981 AGAAGGTTCTTGAGAGCGGTGCAACTGCTTGAGGAGATCCATATATCGGCAGAG 2040
QY 299 PheProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAla 318
DB 2041 TACACTGAGCTTCACCAAGGTCAATGCTTGTGTAACTTACCAAGTCTCGGAATTCCT 2100
QY 319 PheValAsnGluThrSerProLeuSer-----ThrSerGlnLys 331
DB 2101 AAGGCTATAA---ACGGCACCACTCTCCCAATTGTTGATCTACCTGGGCAGACAATCCTACA 2158
QY 332 AlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpSer 351
DB 2159 GCGACT-----TTGCGGAGTACTTGGGTGTGAGC 2188
QY 352 AspAlaProGlySerThrAlaSerLeuThrLeuValAsnAspLeuValIle 371
DB 2189 TTATAGAGTCTCTACGCAAGCACTCTATACCTG----- 2224
QY 372 ThrAlaProAsnGlyThr---LysTyrValGlyAsnAsp----- 383
DB 2225 ACATTCGAGTGGCACATTAACTAGTACGTAGGGACACGAGTACAGAACTTTTGTAGATCT 2284
QY 384 --PheThrAlaProTyrAspAsnAsnTrpAspGly-----ArgAsnAsnV 398
DB 2285 ATGCAACTGAGCCATGATTAAGCCTTTTCTCAGTGAAGTGAATCTTAGAGACATA 2344
QY 398 alGluAsnValPhe-----IleAsnAlaProGlnSerGlyThrTyr----- 411
DB 2345 CCGAGTTTGCTTAGGGTGAATATGATGATAGAGGTCTTTGAGCCAGGTCTCTATGTTG 2404
QY 412 -----ThrValGluValGlnAlaTyrAsnValP 421
DB 2405 GAAGGATAATCATTGATGATCCAAACAGCCAGTTATTGAAGACGAGATCTTGACACAA 2464
QY 421 roValSerProGlnThrPheSer 428
DB 2465 TTGTTATCCGAGAGTTCACT 2487

RESULT 12

US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLERWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1

; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113

Alignment Scores:
Pred. No.: 2,83e-22 Length: 1560
Score: 297.00 Matches: 116
Percent Similarity: 39.95% Conservative: 51
Best Local Similarity: 27.75% Mismatches: 161
Query Match: 13.19% Indels: 90
DB: 15 Gaps: 19

US-09-985-689A-7 (1-433) x US-10-084-846A-113 (1-1560)

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 451 GGGCAGGAGTGCAGCGGTAGCTCATCGACACCGCGCTC-----CGC 492
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
DB 493 ATCACCACACGCACTTCGGCGCGCGCTCTACGGCTACGACGCGCATCGACCAACGAC 552
QY 60 AsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
DB 553 AACACCGCCGAGGACGCGCACGCGCACGCGTGGCCGCGCAGTCCCGGCAAC 612
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
DB 613 GCC-----TACGGCGTCCCAAGAGCAAGATCGTAGCGTCCGCGTCTCTCAACAAC 666
QY 100 GlyGly-----GlyLeuGlyGlyLeuProAlaAsnLeu 110
DB 667 TCCGCCACGAGGACACCGCCAGGTGTCGCCCGCATCGACTGGTGGTCCCGGAAC--- 723
QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
DB 724 -----GCCGTCAAGCGCGCGTCCGCAACATGTCCTCGGC 759
QY 131 AlaProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLys 150
DB 760 GCGCGCGCGACACGCGCTCGACACGCGCGCTAGCAAC-----GCCATGCCC 807
QY 151 AsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyProGlySerGlyThrIleSer 170
DB 808 TCCGCGCTCACTTCGCGCGCGCGCGCAACGAGTGCACCAACGCTCCACGAGGTCA 867
QY 171 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArg 188
DB 868 ---CCGCGACGCGTCAACGAGGCCATCACGCTCGCGCGAGCACCGAGTCCGACGCCAAG 924
QY 189 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 208
DB 925 GCGCGCTACTCAACTAGCGTCCGCTCCTC----- 954
QY 209 ProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 228
DB 955 -----GACCTTCGCGCGCGTTCGTCCATCATCCTCG 987
QY 229 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 248
DB 988 GCC-----TGAACCTCAAGCGACTCGCGGACCAAC 1017
QY 249 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 268
DB 1018 ACCATCTCGGTACGTCTGATGGACCGCGCACGTCGCGGCGCGCGCGCTC----- 1071
QY 269 GluHisPheValLysAsnArgGlyValThrProLys-----ProSerLeuLeuLys 285
DB 1072 ---CACCTCGCGCGCAACCCCTCGCGCACCCCGTCCAGGTCCGACGCGGTGACGTC 1128
QY 286 AlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGly 305
DB 1129 GCGGCCACCAACCGCGGTGCTGCACCAACCCCGCACGGGCTCGCCCAACCGGGTCTCTGTAC 1188


```

QY      306   TtpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGIUthrSerPro 325
DB      1189 GTCCGCCGG---CGGCACGACCCTCCGGCGCGCGCITCGAGAAC---ACGGGTGAC 1242
QY      326   LeuSerThrSerGlnIyAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLys 345
DB      1243 TCACGATCAGGCACAACATCCACGTCGAG-----TCCCCGGTGACG 1284
QY      346   IleSerLeuValTrpSerAspAlaProGlySerThrThraSerLeu----- 361
DB      1285 GTCCTCGCGGTCTCCGGCNACGCGCCCTCGGCCCTCGCGTAGAGGTCCACATCGTCCAC 1344
QY      362   ThrLeuValAsnAspLeuValleThrAlaProAsnGlyThrLysTyrValGly 381
DB      1345 ACGTACATCGGACCTCCAGGTCCAGTGTATCGCCCGCACGCGCATCACGCTC 1404
QY      382   AsnAspPheThrAlaproTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401
DB      1405 AAGTCGTAC-----GGCACC GGCGCAGTTCCGACNACATCAACACCACG 1449
QY      402   Phe---IleAsnAlaProGln-----SerGlyThrTyrThrValGluVal 415
DB      1450 TACTCGGTGAACGCTCTCTCGAGCGCGCAACGCGACGTGAAACTCGGGGTG 1503

RESULT 13
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

Alignment Scores:
Pred. No.:          5,38e-20           Length:       59816
Score:             297.00              Matches:       116
Percent Similarity: 39.95%             Conservative:    51
Best Local Similarity: 27.75%            Mismatches:     161
Query Match:        13.19%              Indels:         90
DB:                  15                 Gaps:          19

US-09-985-689A-7 (1-433) x US-10-084-846A-1 (1-59816)
QY      21   GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB      56648 GGGCAGGAGTGTCGCGGTAGTTCATCGACACCGCGGTC-----CGC 56689
QY      41   SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
DB      56690 ATCACCCACAGCGACTTCGGCGGCGCGGCGCTCTACGGTATCGACGCCATCGACACGAC 56749
QY      60   AsnAsnAlaAsnApproAsnGlyHisGlyThrHisValalaclySerValleuGlyAsn 79
DB      56750 AACACCGCCAGGACGCGCCACGCGCACCGCAGCAGTGGCCGCGCAGCTCCCGCGCAAC 56809
QY      80   AlatrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
DB      1189 GTCCGCCGG---CGGCACGACCCTCCGGCGCGCGCITCGAGAAC---ACGGGTGAC 1242

```

```

56810 GGC-----TACGGCGTCCCAAGAGCCCAAGATCTGTAGCGGTCCGGTGCTGAAACAAC 56866
QY 100 GlyGly-----GlyLeuGlyGlyLeuProAlaAsnLeu 110
Db 56864 TCCGGCCAGCGCACCAACGCCCGCAGGTCTGTCGCCGGCATCGACTGGTCCGCCGGAAC--- 56920
QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 56921 -----GCCGTCGAAGCCGCCGCTGCGCCAAACATGTCCTCCGCG 56956
QY 131 AlaProValAsnGlyAlaTyrThrThrAspSerArgAsnValaAspTyrValArgLys 150
Db 56957 GCGCGCGCCGACACGCGCCTCGACACGCGCGGTACGCAAC-----GCCATGGCC 57004
QY 151 AsnAspMetThrIleLeuPheAlaGlyAsnGlyProGlySerGlyThrIleSer 170
Db 57005 TCCGGGTACACCTTCGCGCGCGCGCAACAGAGTCGACCAACGCTCCACGAGGTCA 57064
QY 171 AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArg 188
Db 57065 ---CCGCGACGGTCACCGAGGCCATCAGGTCGCGCGGACGACCAAGCTCGAGCCCAAG 57121
QY 189 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 208
Db 57122 GCGCGTACTCCAACATACGGCTCCGTCCTC----- 57151
QY 209 ProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer 228
Db 57152 -----GACCTCTTCGCCCCCGGTCGTCATCATCACCTCG 57184
QY 229 AlaArgSerSerLeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAla 248
Db 57185 GCC-----TGGACTCAAGCGACTCGCGCGACCAAC 57214
QY 249 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 268
Db 57215 ACCATCTCCGCTACGTGATGGGACCCCGACGTGGCGGGCGCGCGCGCTC----- 57268
QY 269 GluHisPheValLysAsnArgGlyValThrProLys-----ProSerLeuLeuLys 285
Db 57269 --CACCTCGCGCCCAACCCCTCGGCCACCCCGCTCCAGGTCCGACGCGCTGACGTC 57325
QY 286 AlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGly 305
Db 57326 GCCGCCACCAACCGCGTGTGTACCAACCCCGGACGCGGTCTGCCCAACCGGTCCTGTAT 57385
QY 306 TrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheValAsnGluThrSerPro 325
Db 57386 CTCGGCCGG---CGGCACGACCAACCCCTCGCGGCCCGCGCTTCGAGAAC---ACCGGTGAC 57439
QY 326 LeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLys 345
Db 57440 TACACGATACGACGACAACTCCACGGTCGAG-----TCCCGGTGACG 57488
QY 346 IleSerLeuValTyrSerAspAlaProGlySerThrThrAlaSerLeu----- 361
Db 57482 GTCTCCGGCGTCTCCGCAACGCGCCTCGGCCCTCGCGGTAGAGTCCACATCGCTCCAC 57544
QY 362 ThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381
Db 57542 ACGTACATCGGCGACCTCCAGGTCGAGTCCGCTGATCGCCCCGACGCGCTACACGCTC 57600
QY 382 AsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnAsnValGluAsnVal 401
Db 57602 AAGTCGTAC-----GGCACCGCGCGCGAGTTCGGACACATCATCAACACCCAGC 57644
QY 402 Phe-----IleAsnAlaProGln-----SerGlyThrTyrThrValGluVal 415
Db 57647 TACTCGGTGAACGCTCTCTCGAGGCGGCCAAAGCGACGTGGAAACTCGCGGGTG 57700

RESULT 14
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A

```

RESULT 14
US-10-084
; Sequenc

; Publication NO. US20040006026A1			
; GENERAL INFORMATION:			
; APPLICANT: WEITNAUER, GABRIELE			
; APPLICANT: MUHLENWEG, AGNES			
; APPLICANT: TREFZER, AXEL			
; APPLICANT: BECHTHOLD, ANDREAS			
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES			
; FILE REFERENCE: 1974-005			
; CURRENT APPLICATION NUMBER: US/10/084,846A			
; CURRENT FILING DATE: 2003-02-25			
; PRIOR APPLICATION NUMBER: PCT/EP01/09815			
; PRIOR FILING DATE: 2001-08-24			
; PRIOR APPLICATION NUMBER: DE 101 09 166.4			
; PRIOR FILING DATE: 2001-02-25			
; NUMBER OF SEQ ID NOS: 120			
; SOFTWARE: Patentin Ver. 3.2			
; SEQ ID NO 2			
; LENGTH: 59816			
; TYPE: DNA			
; ORGANISM: Streptomyces viridochromogenes			
US-10-084-846A-2			
Alignment Scores:			
Pred. No.:	5.38e-20	Length:	59816
Score:	297.00	Matches:	116
Percent Similarity:	39.95%	Conservative:	51
Best Local Similarity:	27.75%	Mismatches:	161
Query Match:	13.19%	Indels:	90
DB:	15	Gaps:	19
US-09-985-689A-7 (1-433) x US-10-084-846A-2 (1-59816)			
QY	21	GlyGlnGlyGlnIleValAlaValAlaAAspThrGlyLeuAAspThrGlyArgAsnAAspSer	40
DB	3169	GGGCAGGAGTGACGGCGTACGTCATCGACACGGCGTC-----CGC	3128
QY	41	SerMethHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr	59
DB	3127	ATCACCCACAGCGACTTCGCGCGCGCGGCCTCCTACGGCTACGACGCTCGACACGAC	3068
QY	60	AsnAsnAlaAsnAAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn	79
DB	3067	AACACGCCCAGAGCGGCCACGGCCACGGACCGACGTCGTGCGCGGACGCTCGCGGCAAC	3008
QY	80	AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	99
DB	3007	GCC-----TACGCGCTCGCAAGAGGCCAAGATCGTAGCGCTCCGCTGGAACAAC	2954
QY	100	GlyGly-----GlyLeuGlyGlyLeuProAlaAsnLeu	110
DB	2953	TCCGGCCAGGCGCACCCGCGCAGTCTGTCGCGGCATCGACTGGTTCGCCGGAAC---	2897
QY	111	GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly	130
DB	2896	-----GCCGTCAAGCGCGCGCTCGGCAACATGCTCCTCGGC	2861
QY	131	AlaProValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLys	150
DB	2860	GCGCGCGCCGACACAGCGCCCTCGACACGGCCGTACGCAAC-----GCCATGCC	2813
QY	151	AsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSer	170
DB	2812	TCCGGCGTCACCTTCGCGGTGGCGCGCGGACACGAGTCGACCAACGCTCCACGAGGTCA	2753
QY	171	AlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArg	189
DB	2752	---CCCGCACCGCTCACCGAGGCCCATCACGGTCGCGCGACGACCGCTCGGACGCCAAG	2696
QY	189	ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly	208
DB	2695	GCGGCTACTCCAACTACGGTCGCTCCTC-----	2666
QY	209	ProThrArgAspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSer	228

```
US-09-985-689A-7 (1-433) x US-09-927-827-29 (1-3743)
QY 21 GlyGlnGlnIleValAlaValAlaAspThrGlyLeu-----AspThrGly 36
Db 1502 GGCAGCGCAGCGTGGGGGGTATTGATACCGGCATCACAGTATCGCGACCTCAAC 1561
QY 37 ArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuValAlaLeu 56
Db 1562 GCCAACATCTCGCGGCTACGACTTCATCAGCGATGCGACCAACCGCGATGCAAC 1621
QY 57 GlyArgThrAsnAsnAlaAsp----- 64
Db 1622 GCGCGTACAGCAACCGCGCGAGAGGCGACTGTGTACCGCCCAAGATCGCGCGC 1681
QY 65 -----ProAsnGlyHisGlyThrHisValAlaGlySerValLeuGly 78
Db 1682 GGCATTCCCGCGCGCAGCTCCAGCTGCGCAGCCATGTGGCGCGCAGCTCGCGGCA 1741
QY 79 AsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal----- 92
Db 1742 GTGACCAACAACACACCAGCGGTAGCGCGCACCGCTTACCGCGCCAAGTCTGACCGGTG 1801
QY 93 -----PheGlnSerIleMetAsp----- 98
Db 1802 CGCGTCTCGCAAGTGGGTGGTCTGTCGGATATCGCGAGCCATCGTCTGGGC 1861
QY 99 SerGlyGlyLeuGlyGlyLeuProAlaAsnLeuGln-----ThrLeuPheSerGln 116
Db 1862 TCCGGCGCACCGTACCGCGCATCCCGCCAATGCTAACCGCGCGAGGTGATCAACATG 1921
QY 117 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyVala 136
Db 1922 TCGTCTGGCGCGCGGTAGTGTCTCGACCACTGACAGCGCC-----ATCAACGGTGG 1978
QY 137 TyrThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeu 156
Db 1979 GTGTGCGCGCGCAC-----ACGGTGGTG 2002
QY 157 PheAlaAlaGlyAsnGluGlyProGly---SerGlyThrIleSerAlaProGlyThrAla 175
Db 2003 GTCGCGCGCGCGCAAGATGCTGCTCATGTGTCGGT-----TCGCTCGCGCGCACTGC 2056
QY 176 LysAsnAlaIleThrValGlyAlaThrGluAsn-----LeuArgProSerPheGlySer 193
Db 2057 GCGAACGTGATTGCGTGGCGGCCACCACTCGCGCGCGCGGCGAAGCGCAGTATTCCAAC 2116
QY 194 TyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgAspGly 213
Db 2117 TTCGGCACCGGTATC----- 2131
QY 214 ArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSer--- 232
Db 2132 -----GATGTGCGCGCGCGCGTCTGTCGATCTCTGTCACGCTCAACAGCGGC 2179
QY 233 ---LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 251
Db 2180 ACCACACCGCGGGTAGC-----GCCAGCTATGCTCTCTACAC 2218
QY 252 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 271
Db 2219 GGCACCTCGATGCGTCCGCGCATGTGGCGCGCGTGGTTCGCGCTGTCGTCAGTCG----- 2272
QY 272 ValLysAsnArgGlyValThrProLys-----ProSerLeuLeuLys-----AlaAla 287
Db 2273 GTCGCGCGCGCGCTGACCGCGCAGCGGTGGAAACCTTGTGAAAGAACACCGCGCGT 2332
QY 288 LeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe----- 299
Db 2333 GCTTTACCGCGCGCTGCTCGCGCGGTGCGGTGCGCGCATGCTCAACCGCGATCGCGCG 2392
QY 300 -----ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312
```

Search completed: April 5, 2004, 04:52:56
Job time : 7355.25 secs

```
Db 2393 GTCACCTCGGCGCATCAATGCGGAGCGCGCGGTGGTGGTGAACACACCTTGACC 2452
QY 313 LysSerLeuAsnValAlaPheValAsnGluThrSerProLeuSerThrSerGlnLysAla 332
Db 2453 AACGGCACTCCGGTGACCGGCTGGC-----GGCGGCACTGGCGCGGAATTG 2500
QY 333 ThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp 352
Db 2501 AACTACACCATCACCCTGCGCGCGCGCGCGC-----ACCTTGACGGTGACCAACC 2551
QY 353 AlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeuValIle--- 371
Db 2552 AGCGCGCGCAGC-----GGCGATGCGCGACCTGTATGTGCGC 2587
QY 372 -----ThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAsp 389
Db 2588 GCGCGCACTGACCGACCGCACTCGGCTTACACC-----TGCGCGCCATACCGC 2635
QY 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
Db 2636 AGC-----GGCAATGCGGAGACCTGCACCATCACCGCACCG---TCGGGA 2677
QY 410 ThrTyrThrValGluValGlnAlaTyrAsnValProValSerProGlnThrPheSer 428
Db 2678 ACGTATTACGTGCGTCTGAAGCGCTACAGC-----ACGTTCTCC 2716
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:40:05 ; Search time 2274.58 Seconds
(without alignments)
5684.703 Million call updates/sec

Title: US-09-985-689A-7
Perfect score: 2252
Sequence: 1 NDVARGIVKADVAQNNFGLY.....EVQAYNVSPQTFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q=/cgn2_1/USPTC.spool/US09985689/runat_31032004_161807_4168/app_query.fasta_1.3498
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITG=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=sco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 @CNC_1_1_7084@runat_31032004_161807_4168 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAIDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_estl.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_ges_hum.*
18: em_ges_inv.*
19: em_ges_pln.*
20: em_ges_vrt.*
21: em_ges_fun.*
22: em_ges_mam.*
23: em_ges_mus.*
24: em_ges_pro.*
25: em_ges_rod.*
26: em_ges_pbg.*
27: em_ges_vrl.*
28: gb_gesl.*

29: gb_gses2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the Score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	242	10.7	1605	13	BQ622771	BQ622771 CC Contig
2	235.5	10.5	640	12	BQ395336	BQ395336 BJ395336
3	228	10.1	4198	11	AK029048	AK029048 Mus muscu
4	212	9.4	594	12	BJ393752	BJ393752 BJ393752
5	209.5	9.3	532	29	TA319010P	AL492464 T. brucei
6	205.5	9.1	771	14	CA320325	CA320325 UI-M-FW0-
C 7	197	8.7	508	28	AO652212	AO652212 Sheared D
C 8	197	8.7	2141	13	BQ142519	BQ142519 Contig6 M
9	196	8.7	633	12	BJ369190	BJ369190 BJ369190
10	190.5	8.5	601	12	BJ387574	BJ387574 BJ387574
11	189.5	8.4	665	13	BQ770462	BQ770462 UI-M-F10-
12	186.5	8.3	675	14	CF727824	CF727824 UI-M-HB0-
13	186	8.3	574	29	TA315H10P	AL490202 T. brucei
14	182	8.1	4662	11	BC060527	BC060527 Mus muscu
15	181	8.0	3091	12	BC011375	BC011375 Mus muscu
16	180.5	8.0	718	12	BJ750157	BJ750157 EG02_1090
17	180.5	8.0	1002	29	CNS0606B	AL393417 T3 end of
18	179.5	7.9	895	13	BQ216158	BQ216158 AGENCOURT
19	178.5	7.9	641	12	BJ393925	BJ393925 BJ393925
C 20	178	7.9	530	29	CNS010FO	AL153820 Anopheles
C 21	176.5	7.8	573	14	CA917626	CA917626 sav42b10.
C 22	175.5	7.8	716	28	BZ893395	BZ893395 HL2_0177
C 23	174	7.7	616	14	CF138007	CF138007 UI-EP-BN0
C 24	173.5	7.7	2121	28	BZ424995	BZ424995 100023066
25	172	7.6	614	9	AJ273402	AJ273402 AJ273402
26	172	7.6	681	14	CB690041	CB690041 CEST-54-B
27	171	7.6	449	12	BJ359939	BJ359939 BJ359939
28	170.5	7.6	576	14	CD295943	CD295943 StrPu691.
29	170.5	7.6	594	14	CF846026	CF846026 psHB0362A
C 30	170.5	7.6	1572	11	AY107161	AY107161 Zea mays
31	170	7.5	545	13	BUS75479	BUS75479 TGEStz98
32	169	7.5	712	14	CD311344	CD311344 StrPu691.
33	169	7.5	930	13	BX390734	BX390734 BX390734
C 34	168	7.5	564	28	AO651427	AO651427 Sheared D
35	167	7.4	650	9	AJ274038	AJ274038 AJ274038
36	166	7.4	937	12	BG246418	BG246418 602360428
37	165.5	7.3	688	10	BF636040	BF636040 NF070C09D
38	165.5	7.3	887	13	BQ879057	BQ879057 AGENCOURT
39	165	7.3	601	12	BU365857	BU365857 BJ365857
40	165	7.3	732	13	BU028438	BU028438 QHH14A24.
41	165	7.3	794	14	CF737198	CF737198 UI-M-HD0-
42	164.5	7.3	989	29	CC723222	CC723222 OGLAB20TV
C 43	164	7.3	655	28	BZ349201	BZ349201 hg88c10.g
44	163	7.2	786	13	BUT09126	BUT09126 UI-M-EW0-
45	163	7.2	847	14	CF726182	CF726182 UI-M-GZ0-

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION BQ622771 1605 bp mRNA linear EST 01-JUL-2002
CC Contig67 Conidiobolus cornatus ARSEP 512 Conidiobolus coronatus
CDNA, mRNA sequence.
ACCESSION BQ622771.1 GI:21649940
VERSION BQ622771
KEYWORDS EST
SOURCE Conidiobolus coronatus (Delacroixia coronata)
ORGANISM Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE 1 (bases 1 to 1605)

QY 38 -----AsnAspSer-----SerMethHisGluAlaPheArgGly 48
Db 82 TTTCTTTTCAGATTCAAGTATCCAAATACCAATTAATCAAGTCAATGAAATCATAGAAA 141
QY 49 LysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspPro-----Asn 66
Db 142 GTTGTAACATTATAT-----ACTTACCATGACAATGAAGATTATGTAAT 186
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsn----- 79
Db 187 GTCATGGTACACATGTTTGTGGCTCTGCAGCAGGTACTCCAGAGGATTCCTCATGGGCT 246
QY 80 ----AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 98
Db 247 ATTTCATCATTTAGTGTCTTGCACCTGACGAAAGATTGCAATTTATGAT---CTTCA 303
QY 99 SerGlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTy 118
Db 304 TCTGGAGTCTCAACCAACACACCACCGAAGATTACAGTCAATGTACAAACCATATAT 363
QY 119 SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAla-----ProValAsn 134
Db 364 GATCGAGGTGCAAGACATACATGTTGATTTCTGGGTCTCTGATCTTTGCAAGGTTATAT 423
QY 135 GlyAlaTyThrThrAspSerArgAsnValAspAspTyValArgLys---AsnAspMet 153
Db 424 GFGGTTATCCGATGATGCTGGTGTATGATGCAATCTCTATGAGTACCAAGATTC 483
QY 154 ThrIleLeuPheAlaAlaGlyAsnGlyProGlySerGlyThrIleSerAlaProGly 173
Db 484 TCTATACTAGAGCTGCTGGTAAT---AACGAGCTATTGTCATCTTTATAGCTCAAGCA 540
QY 174 ThrAlaLysAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193
Db 541 ACAGCTAAATGCAATACATGTTGCTGCTGACAAACAGCTCATGTAATATGATGTC 600
QY 194 -----TyrAlaAspAsnIleAsn 199
Db 601 GATGATTCGATATATGATTTCTCAGATAATGCTAAT 639
RESULT 3
AK029048
LOCUS
DEFINITION
AK029048 4198 bp mRNA linear HTC 18-SEP-2003
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
library, clone:473248M11 product:membrane-bound transcription
factor protease, site 1, full insert sequence.
AK029048
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
MEDLINE
PUBMED
10349636
REFERENCE
2
AUTHORS
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE
PUBMED
11042159
REFERENCE
3
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yanamoto, S., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE
PUBMED
11078661
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
(bases 1 to 4198)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saiki, D., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
FEATURES
source
Location/Qualifiers
1..4198
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:473248M11"
/db_xref="MGI:2351080"
/db_xref="taxon:10090"
/clone="473248M11"
/clone_type="skin"
/tissue="skin"
/cDNA_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
417..3575
/note="unnamed protein product; membrane-bound
transcription factor protease, site 1 (MGD|MGI:1927235,
GB|NM_019709, evidence: BLASTN, 99%, match=3782)
putative"
/codon_start=1
/protein_id="BAC36263.1"
/db_xref="GI:26325018"
/translation="MKLVSTWLVLLVGLGRHLGDRGLGTRALEKAPCPSCSHLTK
VERSTVVEYEVIVAFNGYFTAKARNSFISSALKSVEENWR11PRNNPSQSDYSDP
VIQKEQKAGLLIEDHPNFKVTYVPRKFSRLKPAESNPVPCNTRWQKQSSR
TLKASLGSFGFWHTRHRSRLRLRAIPQVATQADVLWQMGYTCANRVAVFD
PGLSEKHPFKVKTWNTNRTLDGLGRTFVAGVIASMRCCQGPAPDAELHIFR
VFTNNQVSYTSWFLDAFNAYIALKKQMDNLNLISIGGDFMDHFPFVDKVMELTANNVIMS

polyA_signal
polyA_site

ORIGIN

```

Alignment Scores:
Pred. No.: 5,91e-11 Length: 4198
Score: 228.00 Matches: 119
Percent Similarity: 38.46% Conservative: 51
Best Local Similarity: 26.12% Mismatches: 166
Query Match: 10.12% Indels: 106
DB: 11 Gaps: 23

US-09-395-689A-7 (1-433) x AK029048 (1-4198)

Qy      8 ValLysAlaAspValAlaGlnAsnAsnPhcGlyLeuTyGlyGlnGlyGlnIleValAla 27
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1005 CTCGAGGCACATGTGCTGTGG---CAGATCGGATACACAGTCTTAATGTCAGAGTTGCT 1061

Qy      28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1062 GTATTTTCATCTCGGCTC-----AGTCAGAAGCATCGCGATTTTAAAG 1103

```

US-09-985-689A-7 (1-433) x AK029048 (1-4198)

Qy	8	ValIysAlaSpValAlaGlnAsnAspPheGlyLeuTyGlyGlnGlyGlnIleValAla	27
Db	1005	CTGCGGCAGATGCTGTGG---CAGATGGGATACACAGTGCTAATGTCAGAGTTGCT	1061
Qy	28	ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg	47
Db	1062	GTTTTTCATTACTGGCTC-----AGTCAGAGGATCCGCATTTAAG	1103
Qy	48	GlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-----	63
Db	1104	AAT-----GTCAAGAGAGAACCAACTGGACCATTAGCGGACCCCTG	1145
Qy	64	---AspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaThrAsn	82
Db	1146	GATCATGGCTTAGGCCATTGGTCAGAGTGTGATTCGACATGAGGGAGTGC	1205
Qy	83	LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGlyGly	101
Db	1206	CAAGATTGCTCCAGATGCAGAGCTGCACATCTTCAGGGCTTTTCAACAACAT----	1259
Qy	102	GlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGly	121
Db	1260	-----CAGGTGCTCATACACTCTGGTTCTCGGATGCTTCAACTATGCC	1304
Qy	122	AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSer	141
Db	1305	ATCCTAAAGAAGATGGACGCTTCAACCTTAGCATCGTGCGCCGACTTCATGGATCAT	1364
Qy	142	ArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAlaAla	159
Db	1365	CCGTTTGTGTGACAGGTGTGGGAATTACACGCTAACATGTAATTATGTGTTCTGCTATT	1424
Qy	160	GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179
Db	1425	GGCAATGATGGACCTCTCTATGGCACTCTGAATACCCCTGCTGATCAGATGGATGTGATT	1484
Qy	180	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn	199
Db	1485	GGAGTGGGTGGCAATGC-----TTTGAAGATTAACATC---	1517
Qy	200	HisValAlaGlnPheSerSerArgGly-----ProThrArgAspGly	213
Db	1518	-----GTCGCTTTTCTTCAGGGGAATGACTACCTCGGAATACCAAGGAGGCTATGCT	1571

Qy	214	ArgIleYsProAspValMetalAlaProGlyThrTyrIleLeuSerAlaAargSerLeu	233
Db	1572	CGFTGGAAAGCCTGACATTGTG-----ACCTATGGTCGTAGATGC GG GTTCGGT	1522
Qy	234	AlaProAspSerSerPheTrpAlaAenHisAspSerLysTyrAlaTyrMetGlyGlyThr	283
Db	1623	GTCAAAAGGGGGCTGC-----CGTGCACTCTCCAGGACC	1655
Qy	254	SerMetAlaThrProIleValAlaGlyAenValAlaGlnLeuArgGluHisPheValLys	273
Db	1656	AGTGTGCTTCCCAGTGGTCTGGGGCGGTCACTTTAGTAGACACAGTACAGAAG	1715
Qy	274	AsnArgGlyValThrProLysProSerLysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu	293
Db	1716	CGGGAGCTGGTG-----AATCTGCCAGTGTGAAGCAAGCTTTGATAGCGTCAGGCCCGG	1769
Qy	294	AspValGlyLeuGlyPheProAenGlyAen-----GlnGlyTrpGlyAArgValThr	310
Db	1770	AGACTT-----CCTGGGTCACATGTTTCGACAGGTCTCATGCCAAGTTGGAT	1817
Qy	311	LeuAspLysSerLeuAenValAlaPheValAenGlnUthrSerProLeuSerThrSerGln	330
Db	1818	CTGCTGCAGCTTATCAGATC-----CTCAGC-----	1844
Qy	331	LysAlaThrTyrSerPheThrAlaGlnAlaGly---LysProLeuLysIleSerLeuVal	349
Db	1845	-----AGCTATAAACCGCAGGCAAGCCTGAGTCTTAGCTACATCGACCTGACT	1892
Qy	350	-----TrpSerAspAlaProGlySerThrThrAlaSerLeuThr	362
Db	1893	GAGTGTCCCTACATGTGGCCCTACTGCTCCAGCCTTACTACTATGAGGAAATGCCAAC	1952
Qy	363	ValValAenAspLeuAspLeuValIleThrAlaProAenGlyThrLysTyrValGlyAen	382
Db	1953	ATCGTTTAAT-----GTCACTCTCAATGGCATGGGCGTCACAGGAAGA	1997
Qy	383	AspPheThrAlaProTyrAspAenAsnTrp-----AspGlyAArgAsnAsn	397
Db	1998	ATTGTGATGAAGCCT-----GAGTGGCGACCCCTATTACACAGAATGGAGACAAC	2048
Qy	398	ValGluAenValPhe-----IleAenAlaProGlnSerGlyThrTyrThrVal	413
Db	2049	ATTGAAGTGGCCCTCTCTACTCTCAGTGTGTGGCCCTGGTCAGGTACCTTGCCATC	2108
Qy	414	GluVal 415	
Db	2109	TCCATT 2114	
RESULT 4			
LOCUS	BJ393752	594 bp mRNA linear EST 08-MAR-2002	
DEFINITION	BJ393752 Dictyostelium discoideum cDNA library, SF Dictyostellum		
ACCESSION	BJ393752		
VERSION	BJ393752.1	GI:19304838	
KEYWORDS	EST.		
SOURCE	Dictyostelium discoideum		
ORGANISM	Dictyostelium discoideum		
REFERENCE	Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellum.		
AUTHORS	i (bases 1 to 594)		
TITLE	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.		
JOURNAL	Full length cDNA of Dictyostelium discoideum at the slug stage		
COMMENT	Unpublished (2002) Contact: Tadasi Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@gene.nig.ac.jp. Location/Qualifiers 1..594 /organism="Dictyostelium discoideum"		
FEATURES	source		

FEATURES
SOURCE

1. .594
/organism="Dictyostelium discoideum"

```

/mol_type="mRNA"
/db_strain="AX4"
/db_xref="taxon:44689"
/clone="dds32b16"
/sex="mat A"
/dev_stage="Slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SF"

ORIGIN
Alignment Scores:
Pred. No.:      8,61e-11      Length:      594
Score:          212.00      Matches:      54
Percent Similarity: 50.2%      Conservative: 39
Best Local Similarity: 29.1%      Mismatches:  62
Query Match:    9.41%      Indels:      30
DB:             12          Gaps:         8

US-09-985-689A-7 (1-433) x BU393752 (1-594)
QY 49 LysIleThrAlaLeuTy:AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHis 68
DB 21 AAAGTTGTAACCTATTATACCATCAACACGCGAGTAGTGATAAAGTGGATGCAC 80
QY 69 GlyThrHisValAlaGlySerValLeuGlyAsn-----Ala 80
DB 81 GGTACACATATTGTGGTTCTGCAGCAGGTACTCCAGAGGATTCCTCAGTTAATATTTCA 140
QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
DB 141 TCATTAGTGGTCTTGCACTGATGCAAGATTGCATTC-----TTTGATTGGCA 191
QY 101 GlyGlyLeuGlyGlyLeu-----ProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyr 118
DB 192 AGTGTTTCATCAAGTTTGACACCTCCATCGGATTTGAAACAATTATATCAACCATATAT 251
QY 119 SerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAla-----ProValAsn 134
DB 252 GACGAGGTCGCAAGATGTGATTTCTGGGTCTCTATCATAGTAGAGGGGTATACA 311
QY 135 GlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsn---AspMet 153
DB 312 GGTAGTTATCATCAGACACTGCTTCAATTGATGATTTCTTTCACATCCAGATTTC 371
QY 154 ThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGly 173
DB 372 ATCATTTCTTAGAGCTCTGTTACACAGCAATACCTATCATCTACTACTACT---CAATCC 428
QY 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg-----188
DB 429 ACTGCAAGAATGTTATTACCGTTGGTGTCTCATCAACAATTCATGAAAATTATTAACT 488
QY 189 -----ProSerPheGlySerTyrAlaAspAsnIleAsn-----HisValAla 202
DB 489 GATGGTCCAAATATATAAATTCAATCATCTGTGATATAAATCAAGAGTTAATATGT 548
QY 203 GlnPheSerSerArg 207
DB 549 GATTTCGATAGCAGA 563

RESULT 5
TA319G10P      532 bp      DNA      linear      GSS 13-DEC-2000
LOCUS          T. brucei sheared genomic DNA clone 319g10, forward sequence,
DEFINITION     Genomic survey sequence.
ACCESSION     AL492464
VERSION       AL492464.1 GI:11867408
KEYWORDS      GSS.
SOURCE        Trypanosoma brucei
ORGANISM      Trypanosoma brucei
              Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
              Trypanosoma.
REFERENCE     1 (bases 1 to 532)
AUTHORS      Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,

```

Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S.E., Rajandream, M.A. and Barrell, B.G.
 Direct Submission
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk
 Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
 Email: nelsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES
 source
 1..532
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /strain="TREU927"
 /db_xref="taxon:5691"
 /clone="319g10"

ORIGIN
 Alignment Scores: 1,27e-10 Length: 532
 Pred. No.: 209.50 Matches: 61
 Score: 47.06% Conservative: 27
 Percent Similarity: 32.62% Mismatches: 66
 Best Local Similarity: 9.30% Indels: 33
 Query Match: 29 Gaps: 7
 DB: 1

US-09-985-689A-7 (1-433) x TA319G10P (1-532)
 QY 92 ValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly-----104
 DB 2 GTAATGCAAGTATAATGTTGTCCGCGGCGGGAAGATTTCACAGGGTGGGTGCC 61
 QY 105 -----GlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyr 118
 DB 62 CATCCAGTCAGAGCTGTCTCTCCCCACGACGTTACTCAATATTCGTCGCGGTATAT 121
 QY 119 SerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThr 138
 DB 122 GCGCTGGAGCCGCGTGTCTCAAACTCGTGGGGTTTGTCTCCCTCCGAGTATTC 181
 QY 139 ThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIleLeuPhe 157
 DB 182 GCTGTGAAAGGATATGATGATGATTCGAGTAGTAGTATGACGATGGCTACTATCTTC 241
 QY 158 AlaAlaGlyAsn-----GluGlyProGlySerGlyThrIleSerAlaProGlyThr 174
 DB 242 TCCACTGCACACAGTATCCAAAGATGGC-----CTAATGACTCGGTGCTG 286
 QY 175 AlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyr 194
 DB 287 GGTAAAGACCGTATGTCGCGGGGTGCACACAAACGCGT-----TTTGACGCTCG 337
 QY 195 AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArg 214
 DB 338 AAAGAC-----ATTGTTCTCTCGTTTCTTCGTCATGTCCTCAACATACGCGTAGG 388
 QY 215 IleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAla 234
 DB 389 ATGAACCCGATCTTGTGCGTCCGCGGGAAGAGGTGCTGCTCTTCTCTCTGGCAA 448
 QY 235 ProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSer 254
 DB 449 GCATCAGCTAAA-----CAATGTAAAGTGGTGGCCCAAGCGGNGTTCATCG 493


```

Db 906 CGAAGCTACGGTGTGCGCAAAATGCCAGCTCTTTGCTGTCAAGGTTCTTGATGACCG 847
Qy 99 SerGlyGlyLeuGlyLeuProAlaAenLeuInThrLeuPheSerGlnAlaTyr 118
Db 846 GGCAGTGGTTCCTACTCCGGTATCATCAGTGGCAGTCTGTGTCGCCAGGACTCCAA 787
Qy 119 Ser-----AlaGlyAlaArgIleHisThrAenSerTrpGlyAlaProValAenGly 135
Db 786 AGTCGTAACTGCCCAATGGCCACATTCCTCCATGAGTCTGGGA-----GGT 739
Qy 136 AlaTyrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 155
Db 738 GGCTACTCGGCTCGCTCAACAGCGGTCCGCTGCTTTGGTTCAGGTCGTGTCCTT 679
Qy 156 LeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAla 175
Db 678 GCCGTCGCGCTGGCAACAGTAACCGGGATGCCCAAAACACCTCT-----CCGCTTCGAG 622
Qy 176 LysAsnAlaIleThrValGlyAlaThr-----GluAenLeuArgProSerPheGlySer 193
Db 621 CCTACTGCTGCTGACCTGTGGTCCACCTGCGTCAGATGACAGCGCATCTACCTTTTCCAAC 562
Qy 194 TyrAlaAspAsnIleAenHisValAlaGlnPheSerSerArgGlyProThrArgAspGly 213
Db 561 TAC-----TGGATCAAT-----GGCGGCCAACACCATCTCTGGCACC 475
Qy 214 ArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeu 233
Db 555 AGAGTT--GTCGATATCTCGCTCTGCTGTAACCGGCATCTCTCCACC-----511
Qy 234 AlaProAspSerPheThrAlaAenHisAspSerLysTyrAlaTyrMetGlyGlyThr 253
Db 510 -----TGGATCAAT-----GGCGGCCAACACCATCTCTGGCACC 475
Qy 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 273
Db 474 TCCATGGCTACTCCCATATTCGCTGT-----CTTGCTGCTACTCTCAGTGCT 427
Qy 274 AsnArgGlyValThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 293
Db 426 CTCAGCGGCAAGACTAGCCCTCGCGCTCTTGGCAGAGATCCAGGACATCTTACCAC 367
Qy 294 AspValGlyLeuGlyPheProAsnGly 302
Db 366 AACGTGATCCGCAATGTGCGCGCTGGC 340

RESULT 9
BJ369190
LOCUS
DEFINITION
  BJ369190 Dictyostelium discoideum cDNA library, CF Dictyostelium
  discoideum cDNA clone ddc49116 5', mRNA sequence.
ACCESSION
  BJ369190
VERSION
  BJ369190.1 GI:19278573
KEYWORDS
  EST.
SOURCE
  Dictyostelium discoideum
  Organism
  Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE
  1 (bases 1 to 633)
  Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  Full length cDNA of Dictyostelium discoideum at the culmination
  stage
  Unpublished (2002)
  Contact: Tadasu Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshin@genes.nig.ac.jp.
  Location/Qualifiers
  1..633
  /organism="Dictyostelium discoideum"

```

```

/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc49116"
/sex="mat A"
/dev stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Alignment Scores:
Pred. No.: 3.9e-09 Length: 633
Score: 196.00 Matches: 58
Percent Similarity: 41.55% Conservative: 28
Best Local Similarity: 28.02% Mismatches: 49
Query Match: 8.70% Indels: 72
DB: 12 Gaps: 6

US-09-985-689A-7 (1-433) x BJ369190 (1-633)

Qy 135 GlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLys---AenAspMet 153
Db 6 GGTGTTATTCCGATGATGCTGGTGTGTTATGTCATTCCTCTATGAGTACCAGAAATTC 65
Qy 154 ThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGly 173
Db 66 TCTATCTAAGAGCTGCTGTAAT---AACGACTATTGTCATCTTTATAGCTCAAGCA 122
Qy 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAenLeuArgProSerPheGlySer 193
Db 123 ACAGCTAAAAATGCATTTACAGTTGGTGTGTCAGCAACACAGCTCATGTAAATATTGTGTC 182
Qy 194 -----TyrAlaAspAsnIle-----198
Db 183 GATGATTTGGAAATATTATGATTTCTCAGATTAATTTTCAAAGACCATGTTTATTC 242
Qy 198 -----198
Db 243 GATAAGAAGTATTGTAATTATACGACGCTAAATGTGTGTCAGAGGTTTCAAATGTATAA 302
Qy 198 -----198
Db 303 GGTTTACAAATTATGTTGTCAGCATCTATTAAACAAATGCATCGGATTCATTCACAACA 362
Qy 199 -----AenHisValAlaGlnPheSerSerArgGlyProThrArg 211
Db 363 CAACCTCAATTTTATAATGAAATAATATATGGGATCATTTCTCATCAAGGGTCCACACAT 422
Qy 212 AspGlyArgIleLysProAspValMetAlaProGlyThrTyrIleLeuSerAlaArgSer 231
Db 423 GATGTTAGATGGAACCTGATATAGTTGACCTGGTGAATATATTATACATCGGCAAGATCA 482
Qy 232 Ser-----LeuAlaProAspSerSerPheThrAlaAenHisAspSer 245
Db 483 AATGGTGAGAAATTCACACAGCAATGTGGTGGTCTCTTTA-----CCAAATGCCAAT 536
Qy 246 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 265
Db 537 GGTCTAATGTCAATATCTGGTACATCAATGGCAACCACTTGGCAACAGCAGCAACAACA 596
Qy 266 GlnLeuArgGluHisPheVal 272
Db 597 ATTCTTAGACAATATTAGTT 617

RESULT 10
BJ387574
LOCUS
DEFINITION
  BJ387574 Dictyostelium discoideum cDNA library, SF Dictyostelium
  discoideum cDNA clone ddc3a18 5', mRNA sequence.
ACCESSION
  BJ387574
VERSION
  BJ387574.1 GI:19296958
KEYWORDS
  EST.
SOURCE
  Dictyostelium discoideum
  Organism

```

```
REFERENCE
1 (bases 1 to 601)
AUTHORS
Urushibara H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE
Full length cDNA of Dictyostelium discoideum at the slug stage
JOURNAL
Unpublished (2002)
COMMENT
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..601
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds3a18"
/sex="mat A"
/dev_stage="slug stage"
/clone_lib="Dictyostelium discoideum cDNA library, SP"
ORIGIN
Alignment Scores:
Pred. No.: 1.28e-08 Length: 601
Score: 190.50 Matches: 57
Percent Similarity: 44.86% Conservative: 26
Best Local Similarity: 30.81% Mismatches: 63
Query Match: 8.46% Indels: 39
DB: 12 Gaps: 6
US-09-985-689A-7 (1-433) x BJ387574 (1-601)
QY 197 AsnLeuHisValAlaGlnPheSerArgGlyProThrArgAspGlyArgIleLys 216
Db 51 AATGAGAAATATATTGTTCTCATCAAAAGGTCACACATGATGAGTGAATGAAA 110
QY 217 ProAspValMetAlaProGlyThrTyrlleLeuSerAlaArgSerLeuAla ---- 234
Db 111 CCTGATTGTTGCCCTGTTGTAATATTACATCGCAGATCAATGTCGCAATACA 170
QY 235 -----ProAspSerPheTrpAlaAsnHisAspSerLysTyrlaTyMet 250
Db 171 ACAGACCAATGGTGATGCTCTTTA---CCAAATACAAATCATATTGGCG---ATA 224
QY 251 GlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHis 270
Db 225 TCTGGTACATCAATGGCACTCATTTCGACGACGACCAACAATTCCTAGACAATAT 284
QY 271 -----PheValIleAsnArgGlyValThrProLys 280
Db 285 TTAGTTGATGTTATATCCACTGGTTCAATTGTAGATCAATAATAATACCAACT 344
QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
Db 345 GGATCATATTAAAGCATTAATGATTAATTAATGCTCAGTTATTAATGTCATTTCAA 404
QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsn----- 303
Db 405 TTGATTACATCACTAAGTATTACATATCCATCAACCAAGTTTGTGAATTTTCAGGT 464
QY 304 -----GlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db 465 GCAAGTTTAGTTCAAGGTGGGTGCTTAGAATGAGTAAATGCTTACATGTTGTCAAT 524
QY 320 ValAsnGlnThrSerProLeuSerThrSerGlnLysAlaThrTyrlaGln 339
Db 525 AATTAATAGTAATAATAATAAACAACATCAGATGCTTAATAAATTTGTTGTTAT 584
QY 340 AlaGlyLysProLeu 344
Db 585 GGTGATTAGATTTA 599
```

```
RESULT 11
BQ770462
LOCUS
DEFINITION
UI-M-F10-bvY-m-19-0-UI.r1 NIH BMAP_F10 Mus musculus cDNA clone
IMAGE:5702970 5', mRNA sequence.
BQ770462
VERSION
BQ770462.1 GI:21978936
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 665)
NIH-MGC http://mgi.nci.nih.gov/.
REFERENCE
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
```

```
Seq primer: pYX-5.
Location/Qualifiers
1..665
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5702970"
/tissue_type="whole brain"
/dev_stage="embryo 12.5dpc"
/lab_host="DH10B (TI phage resistant)"
/clone_lib="NIH BMAP_F10"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CAGCACACGAC. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
```

```
FEATURES
source
```

ORIGIN

```
Alignment Scores:
Pred. No.: 1.91e-08 Length: 665
Score: 189.50 Matches: 70
Percent Similarity: 43.51% Conservative: 34
Best Local Similarity: 29.29% Mismatches: 94
Query Match: 8.41% Indels: 41
DB: 13 Gaps: 9
```

US-09-985-689A-7 (1-433) x BQ770462 (1-665)

```
QY 63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaThrAsn 82
Db 14 GATGATGGCTAGGCGCATGGCACATTCCTTCAGGTGTGATTCAGCATGAGGAGTGC 73
QY 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerClyGly 101
Db 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerClyGly 101
```

Db 74 CAAGGATTGCTCCAGATGAGAGTGCACATCTTCCAGGCTTTTACCAACAAT----- 127

Qy 102 GlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGly 121

Db 128 -----CAGGTGCTTACACATCTTGGTTCCTGGATGCCCTTCAATATGCC 172

Qy 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAspSer 141

Db 173 ATCCTAAAGAGATGAGCTTCTCAACCTTACATCGGTGGCCGACCTTCATGATCAT 232

Qy 142 ArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAlaAla 159

Db 233 CGCTTTGTTGACAGGTGGGAATTAACAGCTACAAATGAATGATGTTCTGCTATT 292

Qy 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179

Db 293 GGCAATGATGAGACCTCTCATGGCACTCTGAATAACCTGCTGATGATGATGATGAT 352

Qy 180 ThrValGlyAlaThrIleAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199

Db 353 GGAGTGGGTGGCAATTGAC-----TTGAGAGATAACATC--- 385

Qy 200 HisValAlaAlaGlnPheSerSerArgGly-----ProThrArgAspGly 213

Db 386 -----GCTCGCTTTTCTTCCAGGGGAATGACTACCTGGGAATTACCAAGGAGCTATGGT 439

Qy 214 ArgIleLysProAspValMetAlaProGlyThrIleLeuSerAlaArgSerSerLeu 233

Db 440 CGTGTGAGACCTGACATTTGC-----ACCTATGGTGTGAGTGGCGGGTTCCGGT 490

Qy 234 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 253

Db 491 GTGAAAGGGGGGTGTC-----CGTGCACTCTCAGGGACC 523

Qy 254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 273

Db 524 AGTGTGCTTCNCAGGTGCTGCTGGGCGGTCACCTGTTAGTAAAGCACAGTACAGAG 583

Qy 274 AsnArgGlyValThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla 292

Db 584 CGGAGCTGCTGTG-----AATCTGCCAGTGTGAAGCACGCTTTGATAGCTGACGCC 634

RESULT 12
CF727824
LOCUS
DEFINITION UI-M-HB0-ckl-m-24-0-UI-r1 NIH_BMAP_H80 Mus musculus cDNA clone
IMAGE:30550487 5', mRNA sequence.

CF727824
EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 675)
NIH-MGC http://img.ncbi.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousef1.html
This clone was contributed by the Brain Molecular Anatomy Project (BMAP).

Seq primer: pYX-5.
Location/Qualifiers
1..675
/organism="Mus musculus"
/mol_type="mRNA"

FEATURES
source

ORIGIN

Alignment Scores:
Pred. No.: 3,93e-08 Length: 675
Score: 186.50 Matches: 69
Percent Similarity: 43.46% Conservative: 34
Best Local Similarity: 29.11% Mismatches: 93
Query Match: 8.28% Indels: 41
DB: 14 Gaps: 9

US-09-985-689A-7 (1-433) x CF727824 (1-675)

Qy 63 AsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAlaThrAsn 82

Db 25 GATGATGGGTAGGCCATGGCCACATTCGTCAGTGTGATGCCAGCATGAGGAGTGC 84

Qy 83 LysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGlyGly 101

Db 85 CAAGGATTTGCTCCAGATGCGAGCTGCACATCTTCAGGCTCTTTACCAACAAT----- 138

Qy 102 GlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGly 121

Db 139 -----CAGGTGCTTACACATCTTGGTTCGTGATGCTTCACTATGCC 183

Qy 122 AlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAspSer 141

Db 184 ATCCTAAAGAGATGAGCTTCTCAACCTTAGCATCGGTGGCCGACCTTCATGATCAT 243

Qy 142 ArgAsnValAspAspTyr-----ValArgLysAsnAspMetThrIleLeuPheAlaAla 159

Db 244 CGCTTTGTTGACAGGTGCTGGGAATTAACAGCTAACCAATGATATGTTCTGCTATT 303

Qy 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179

Db 304 GGCATGATGAGACCTCTCTATGGCACTCTGAATAACCCCTGCTGATCAGATGATGAT 363

Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199

Db 364 GGAGTGGGTGGCATTGAC-----TTGAGAGATAACATC--- 396

Qy 200 HisValAlaGlnPheSerSerArgGly-----ProThrArgAspGly 213

Db 397 -----GCTCGCTTTTCTTCCAGGGGAATGACTACCTGGGAATTAACAGGAGCTATGGT 450

Qy 214 ArgIleLysProAspValMetAlaProGlyThrIleLeuSerAlaArgSerSerLeu 233

Db 451 CGTGTGAGACCTGACATTTGC-----ACCTATGCTGCTGGAGTGGCGGGTTCCGGT 501

Qy 234 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThr 253

Db 502 GTGAAAGGGGGGTGTC-----CGTGCACTCTCAGGGACC 534

/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30550487"
/tissue_type="whole eye"
/dev_stages="embryo 12.5,13.5,14.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_H80"
/note="Organ: Eye; Vector: pYX-Asc; Site: 1: EcoR I;
Site 2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TTATGTAGT. This library was created for the University
Iowa Brain Anatomy Project (BMAP). Gene Discovery in the
Developing Mouse Nervous System", supported by National
Institute of Mental Health (NIMH)."

Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>; tom-casavant@uiowa.edu
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldio, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A.,
Fishler, K., Koppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K.,
Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J.,
Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6678418
This clone has the following problem: frame shifted.

FEATURES
source
1. 4662
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6844459"
/tissue_type="Brain, mouse, 13.5, 14.5, 16.5, 17.5 dpc"
/clone_lib="NIH_BMAP_FYO"
/lab_host="DH10B"
/note="vector: pYX-ASC"

ORIGIN

Alignment Scores:
Pred. No.: 2,99e-06 Length: 4662
Score: 182.00 Matches: 102
Percent Similarity: 36.54% Conservative: 69
Best Local Similarity: 21.79% Mismatches: 159
Query Match: 8.08% Indels: 138
DB: 11 Gaps: 22

US-09-985-689a-7 (1-433) x BC060627 (1-4662)

QY 10 AlaAspValAlaGlnAsnAspPheGlyLeuTyrglyGlnGlyGlnIleValAlaValAla 29
DB 803 GCTGAGATGCTGAATTCTCTGTAACATTTATGACGATGGAAACCTGCTCTCCATTGTG 862
QY 30 AspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLys 49
DB 863 ACCAGCGA----- 871
QY 50 IleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspProAsnGlyHisGly 69
DB 872 -----GGAGCTCATGGA 883
QY 70 ThrHisValAlaGlySerValLeuGlyAsn-----AlaThrAsnLysGlyMet 85
DB 884 ACCCATGTAGCAAGTATAGCCGAGGCGATTTTCCAGAGAGCCTGGAACGGAATGGAGTT 943
QY 86 AlaProGlnAlaAsnLeuValPhe-----GlnSerIleMetAsp 98
DB 944 GCTCTGTGGTCAAAATTCATCCATTAAAGTGGTGATACACGGCTAAGCACTATGGAA 1003
QY 99 SerGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAla--- 117
DB 1004 ACAGGCACAGGCGCTC-----ATCAGAGCTATGATAGAGTTATTAATCATAGTGTCAT 1057
QY 118 -----TySerAlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsn 134
DB 1058 CITGTCAACTACAGTATGAGAGCAACTCAT-----TGG-----CCAATTCCT 1102
QY 135 GlyAlaTyThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThr 154
DB 1103 GGGAGAATT-----TGTGAAGTAATTAATGAAGCATGTGGAACATTAACAATT 1153
QY 155 IleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaPro---Gly 173
DB 1154 TATGTTCAAGTCTGGAAATATGTCATGCTTCTACAGTGGGTGTGTCACAGGAGA 1213

QY 174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193
DB 1214 ACTACATCCAGTGTGATAGGTGTGGAGCT----- 1243
QY 194 TyrAlaAspAsnIleAsnHisValAlaGlnPhe----- 204
DB 1244 TATGTTCCCTGATATGATGGTTGCAGAGTATTCCTGAGAGAGAACTGCCTGCAAT 1303
QY 205 -----SerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
DB 1304 CAATATACATGGTCTTTAGAGGCCCAAGTCTGATGGAGCCCTCGGTGTGAGCATCAGT 1363
QY 221 AlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
DB 1364 GCACAGGAGGTGCTATTGCTTCTGCTCCTAAC-----TGG 1399
QY 241 AlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 260
DB 1400 ACATTGAGGGGAGCTCAG-----CTAATGAATGGGACATCAATGTCTCCCAATGCC 1453
QY 261 AlaGlyAsnValAla-----GlnLeuArgGluHisPheValLysAsnArgGly 276
DB 1454 TGTGTGGCATTTGCCCTGCTGTTTTCAGGGCTGAAACCAATAATGTTGACTACTGTGA 1513
QY 277 ValThrProLysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGly 296
DB 1514 CACTCAGTCAGAGAGCTCTAGAAATACCTGAATAAAAGCTGACAAATATAGAGTATT 1573
QY 297 LeuGlyPheProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsn 316
DB 1574 GCC-----CAAGGACATGGAATTTATTCAGGTGACAAAGCTTATGAC 1615
QY 317 ValAlaPheValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTySerPhe 336
DB 1616 TACCTCATCAAAAT-----ACATCATTTGCTAACAGATAGGTTTT 1657
QY 337 ThrAlaGlnAlaGlyLys-----ProLeuLysIleSerLeu 348
DB 1658 ACAGTTACTGTTGGAATAAACCGTGTATCTACCTCCGAGATCTCTGTCAGGTGGCTGCT 1717
QY 349 ValTyrSerAsp-----AlaProGlySerThrThrAlaSer 360
DB 1718 CCT---TCAGATCATGTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1774
QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyVal 380
DB 1775 GAAATAATATCTTTTCAGCTTTCATTTAGCTTTAACTTCA----- 1813
QY 381 GlyAspAspPheThrAlaProTyAspAsnAsnTyrAspGlyArgAsnAsnValGluAsn 400
DB 1814 ---AATTCATCTTGGGTCAGTGTCCCGCCACTTTGGAACCTCATGAACTCAGTGTGGCAC 1870
QY 401 ValPheIleAsnAlaPro-----GlnSerGlyThrTyThrValGluValGln 416
DB 1871 ATAAACATACGTTGGAGCCCGAGGGCTTAAAGAGAGGGTTACATTATACAGAGGTATGT 1930
QY 417 AlaTyAsnValProValSerPro 424
DB 1931 GGCTATGATATA---GCATCCCCC 1951
RESULT 15
BC011275 3091 bp mRNA linear HTC 19-NOV-2003
LOCUS Mus musculus tripeptidyl peptidase II, mRNA (cDNA clone
DEFINITION IMAGE:3674891), containing frame-shift errors.
ACCESSION BC011275
VERSION BC011275.1 GI:15030057
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3091)

Db 1584 TACCTCATTCAAAT-----ACATCATTGCTAACAGATTAGGTTTT 1625

Qy 337 ThrAlaGlnAlaGly 341

Db 1626 ACAGTTACTGTTGGA 1640

Search completed: April 4, 2004, 11:55:24
Job time : 2286.58 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:37:54 ; Search time 335.908 Seconds
(without alignments)
5476.111 Million cell updates/sec

Title: US-09-985-689A-7
Perfect score: 2252
Sequence: 1 NDVARGIVKADVAQNNGFLY.....EYQAYNPVSPQTFSLAIVH 433

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPRO.spool/US0985689/runat_31032004_161806_4145/app.query.fasta_1.3498
-DB=N Geneseg 29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=p2n -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689@cgn_1_1238@runat_31032004_161806_4145 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseg 29Jan04:.*
1: Genesegn1980s:.*
2: Genesegn1930s:.*
3: Genesegn2000s:.*
4: Genesegn2001as:.*
5: Genesegn2001bs:.*
6: Genesegn2002as:.*
7: Genesegn2003as:.*
8: Genesegn2003bs:.*
9: Genesegn2003cs:.*
10: Genesegn2004s:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2252	100.0	3003	AAV82382	AAV82382 Bacillus
2	2125.5	94.4	1923	AAx37278	AAx37278 Bacillus
3	2120.5	94.2	1923	AAx37279	AAx37279 Bacillus
4	2110.5	93.7	1920	AAx37277	AAx37277 Bacillus
5	2032	90.2	1299	AAQ27516	AAQ27516 Alkali-pr
6	432.5	19.2	1977	AAAT85667	AAAT85667 Thermococ
7	432.5	19.2	1977	AAx05926	AAx05926 WO9856926
8	396	17.6	1236	AAx05920	AAx05920 Hyperther

9	396	17.6	1566	2	AAAT85668	AAAT85668 Pyrococcus
10	396	17.6	1962	2	AAAT85695	AAAT85695 Pyrococcus
11	396	17.6	1962	2	AAx05929	AAx05929 Hyperther
12	390	17.3	1977	2	AAAT85669	AAAT85669 Protease
13	369.5	16.4	2121	4	ABL54900	ABL54900 T. yonsei
14	327.5	14.5	2539	2	AAAT61454	AAAT61454 Streptomy
15	327.5	14.5	2809	2	AAAT61455	AAAT61455 DhpA-mel
16	325	14.4	1329	6	ABK74643	ABK74643 Bacillus
17	323	14.3	135638	7	ABX34289	ABX34289 S. atrool
18	311.5	13.8	2835	2	AAAT08141	AAAT08141 Hyperther
19	311.5	13.8	4765	2	AAAT08132	AAAT08132 Protease
20	311.5	13.8	4765	2	AAAT85670	AAAT85670 Pyrococcus
21	311.5	13.8	4765	2	AAAT85671	AAAT85671 WO9856926
22	309	13.7	2532	2	AAQ23134	AAQ23134 Encodes R
23	297	13.2	1560	7	ABZ37569	ABZ37569 Streptomy
24	297	13.2	1859	2	AAAT85677	AAAT85677 Thermococ
25	297	13.2	59816	7	ABZ37516	ABZ37516 Streptomy
26	297	13.2	59816	7	ABZ37515	ABZ37515 Streptomy
27	289	12.8	3413	2	AAV72330	AAV72330 F. balust
28	279.5	12.4	3743	9	ADD24901	ADD24901 DNA encod
29	278.5	12.4	546	4	ABL53453	ABL53453 T. yonsei
30	276.5	12.3	3788	9	ADD24905	ADD24905 DNA encod
31	276	12.3	898	2	AAAT08131	AAAT08131 Hyperther
32	275.5	12.2	4716	9	ABQ80437	ABQ80437 CSPa codi
33	275.5	12.2	4740	7	ABZ58957	ABZ58957 Group B S
34	271.5	12.1	1306	6	ABL55784	ABL55784 Bacillus
35	271.5	12.1	1330	6	ABL55787	ABL55787 Bacillus
36	270	12.0	564	2	AAAT08134	AAAT08134 DNA seque
37	270	12.0	564	2	AAAT85676	AAAT85676 Thermococ
38	268.5	11.9	4650	6	ABN71526	ABN71526 Streptoco
39	268.5	11.9	4710	6	ABN71162	ABN71162 Streptoco
40	268.5	11.9	4734	6	ABN70525	ABN70525 Streptoco
41	268.5	11.9	4770	6	ABN69191	ABN69191 Streptoco
42	268.5	11.9	110000	6	ABN71527_19	Continuation (20 o
43	258	11.5	2166	6	ABD34433	ABD34433 C-termina
44	258	11.5	2166	6	ABL40498	ABL40498 DNA const
45	256.5	11.4	807	2	AAQ98740	AAQ98740 DNA encod

ALIGNMENTS

RESULT 1
AAV82382
ID AAV82382 standard; DNA; 3003 BP.

XX AC AAV82382;
XX 12-APR-1999 (first entry)
XX DE Bacillus JP170 protease gene.
XX KW Protease; detergent; surfactant; leather processing; debittering;
XX KW flavour; ss.
XX OS Bacillus sp.
XX FH Key Location/Qualifiers
XX FT CDS 846..2771
FT sig_peptide /*tag= a
FT FT 846..944
FT mat_peptide /*tag= b
FT 1470..2768
FT /*tag= c

1020

XX Sloma A, Christianson L;
 XX WPI; 1999-080908/07.
 DR P-PSDB; AAW89547.
 XX
 XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 PT dishwashing detergents and for leather processing.
 XX
 XX Claim 11; Page 52-53; 77pp; English.
 XX
 XX This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JPI170 (NCIB 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JPI170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50),
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW82410-11), reconstruction of 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins
 XX
 SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4.81e-170 Length: 3003
 Score: 2252.00 Matches: 433
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689a-7 (1-433) x AAW82382 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
 Db 1470 AATGACGTGGCCGTCGGATTGTGAAGCAGACGTCGCACAAAATAACTTTGGCTTATAT 1529
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGACAGGACAGATTGTAGCAGTTCTGATCTGGCTTGTATACAGAGAGAAATGACAGT 1589
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGCATGAAGCATTCGCGGTAAAGATTACCGCAGCTATATGCACTGGCGAGAACGAAT 1649
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
 Db 1650 AACGCCAATGATCCAAATGGACATGGACCCCATGTGCTGGATCTGTGTAGGAATGCT 1709
 QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
 Db 1710 ACAATAAAGGATGGCACCAGCCCAATCTAGTCTTTCAATCTATTATGATAGTGGT 1769
 QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
 Db 1770 GGAGGGCTGGGAGGACTACTCTCTAATCTACAAACATTATTTCAGTCAAGCATATAGTCT 1829
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrThrAsp 140
 Db 1830 GGAGCAGAGATTTCATCAGATTTCATGGGGGCTCCAGTAAGCGTGCCATACGACACAG 1889
 QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
 Db 1890 TCTCGAAATGTTGATGATTATGTGAGAAAAATGATATGACGATTTCTTTTTCGGCCCGGA 1949

QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 1950 AATAGGACACAGGTAGCGGTACAAATCAGTGCACAGCAACAGCAAAATGCGATTACA 2009
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
 Db 2010 GTTGGGGCAACGAAACCTACGTCGAAGCTTCGATCTTATGGGATATATTAAACCAT 2069
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
 Db 2070 GTTGCTCAATCTCTTCACGAGGTCCTACTAGAGATGACGATATTAAAGCCGAGCATG 2129
 QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 Db 2130 GCACAGATGATGATATTTCTCTCTGTAGATCATCATTAGCTCCAGATTCCTCATTTCTGG 2189
 QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 2190 GCAAACCATGATAGTAAATATGCTACATGCGTGGTACTTCTATGGCTACTCCAAATGTA 2249
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
 Db 2250 GCAGTAATGTTGCAAAATTAAGGAGCATTTTGTGAAAAATAGAGGGTAACTCTTAAG 2309
 QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAspValGlyLeuGlyPhePro 300
 Db 2310 CCTTCCCTTTTAAAGCTGCTTTAATTCAGGTGCTGGGATGTTGGACTTGGCTTTCCA 2369
 QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
 Db 2370 AATGGTAACCAAGGATGGGAAGATACGTTAGATAAATCCCTAAATGTGCAATTTGTG 2429
 QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340
 Db 2430 AATGAACAGAGCCCTTTATCAACAAGTCAAAAAGCAACATATTCGTTTACGGCTCAAGCT 2489
 QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360
 Db 2490 GGTAAACCCCTTAAAAATATCACTTTGTTGGTCAGATGCACCAAGTAGCAGCAGGCATCA 2549
 QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380
 Db 2550 CTAACTTTAGTGAATGATTTAGACTTAGCTATCATCTGCACCAATGGAATTAATACGTC 2609
 QY 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
 Db 2610 GGAATGACTTTTACAGCACCGGTATGATACAAATGGGATGGCAGAAACAAACGCTGAAAT 2669
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnVal 420
 Db 2670 GTGTTTATCAATGCTCTCTCAAGCGGAGAGTATACATGCAAGTGCAGGCTTACAATGTA 2729
 QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
 Db 2730 CCAGTAAGTCCGCAAAACCTTTCTTTAGCGATTGTACAT 2768

RESULT 2
 AAX37278
 ID AAX37278 standard; DNA; 1923 BP.
 XX AC AAX37278;
 XX
 XX 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 XX *Bacillus* alkaline protease encoding DNA.
 XX Alkaline protease; *Bacillus*; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX *Bacillus* sp.
 OS
 XX

PN	WO9918218-A1.	Db	979	GCTGGTCCAGAAATTCATACAACTCTCTGGGGAGCAGCTGAATGGGGCTTACACAACA	1038
XX					
PD	15-APR-1999.	Qy	140	AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaA	159
XX					
PF	07-OCT-1998; 98WO-JP004528.	Db	1039	GATTCACAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCTTTTCGTGCC	1098
XX					
PR	07-OCT-1997; 97JP-00274570.	Qy	160	GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179
XX					
PR	(KAOS) KAO CORP.	Db	1099	GGGAATGAAGACCGAAGCGGGAACCATCAGTGCCACCGACAGCTAAATATGCAATA	1158
XX					
PI	Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;	Qy	180	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn	199
PI	Shikata S, Nomura M;	Db	1159	ACACTCGAGCTACGGAAACCTCCGCGCAAGCTTTGGTGCTTATGGGCAATATCAAC	1218
XX					
DR	WPI; 1999-287736/27.	Qy	200	HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal	219
XX					
XX	P-PSDB; AAY17088, AAY17090.	Db	1219	CATGTGGCACAGTTCTCTTCACTGGACCGCAACAGGATGGACGGATCAACACCGATGC	1278
PT	Alkali protease from Bacillus used in washing powders.				
XX					
PS	Disclosure; Page 58-63; 71pp; Japanese.	Qy	220	MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe	239
XX					
CC	The invention relates to alkaline proteases produced by strains of	Db	1279	ATGGCACCGGAACGTTCTACTATCAGCAAGATCTTCTCTTGACCGGATTCCTCTTC	1338
CC	Bacillus. The proteases ability to digest casein is not inhibited by	Qy	240	TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle	259
CC	oleic acid and they have a high stability to oxidising agents. The	Db	1339	TGGCGCAACCATGACAGTAAATATGATATGATGGGTGGAACGTCCTATGGCTACACCGATC	1398
CC	alkaline protease of the invention has the following properties: (a) it				
CC	is active over the pH range 4-13 and has at least 80% of its optimum	Qy	260	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro	279
CC	activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is	Db	1399	GTTGTGGAACGTTGGACACAGCTTCGTGAGCATTTTGTGAAACACAGGAGCATCACACCA	1458
CC	stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)	Qy	280	LysProSerLeuLeuLysAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe	299
CC	its ability to digest casein is not inhibited by oleic acid; (e) it has	Db	1459	AAGCTTCTCTATTAAAGCGGCACTGATTCGCGGTGCGAGCTGACATCGGCTTGGGTAC	1518
CC	molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be	Qy	300	ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe	319
CC	used as enzymes in washing compositions for use in automatic dishwashers	Db	1519	CCGAACGGTAACCAAGGATGGGACGAGTGATGGATTAATCCCTGAACTTGGCTAT	1578
CC	and for washing clothes. The stability to oxidising agents allows the	Qy	320	ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln	339
CC	enzyme to be an effective component of washing compositions including	Db	1579	GTGAACGAGTCCAGTTCTCTATCCACCGCCAAAGACGAGCTACTCGTTTACTGTACT	1638
CC	bleaches. The present sequence represents an alkaline protease encoding	Qy	340	AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla	359
CC	DNA. (Updated on 20-MAR-2003 to correct DR field.)	Db	1639	GCCGGCAAGCCCTTGAATACTCCCTGGTATGGTCTGATGCCCTGGCAGCAACTGCT	1698
XX					
SQ	Sequence 1923 BP; 576 A; 418 C; 475 G; 454 T; 0 U; 0 Other;				
Alignment Scores:					
Pred. No.:	3,476-160	Length:	1923		
Score:	2125.50	Matches:	406		
Percent Similarity:	97.93%	Conservative:	19		
Best Local Similarity:	93.55%	Mismatches:	8		
Query Match:	94.38%	Indels:	1		
DB:	2	Gaps:	1		
US-09-985-689A-7 (1-433) x AAX37278 (1-1923)					
Qy	1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPhedGlyLeuTyr	20			
Db	619 AATGATCTTCGGGTGGAATGTCAAGCGGATGTGGCTCAGACGAGCTACGGGTGTAT	678			
Qy	21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40			
Db	679 GGACAGGACAGATCGTAGCGGTGCCGATACAGGGCTTGATCAGGTGCAATGACAGT	738			
Qy	41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn	60			
Db	739 TCGATGATGAAGCTTCCGCGGAAATTAATCTGCAATATATCATTTGGACCGAGCAAT	798			
Qy	61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--	79			
Db	799 AATGCCAATGATACGAATGTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC	858			
Qy	80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	99			
Db	859 TCCACTAATAAGGAATGGCGCTCCAGCGGAATCTAGTCTTCCAAATCTATCATGATGATG	918			
Qy	100 GlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer	119			
Db	919 GGTGGGGACATGTGAGGACATACCTTCGAATCTGAAACCTTATTCACCAAGCATACAGT	978			
Qy	120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr	139			

RESULT 3
 AAX37279
 ID AAX37279 standard; DNA; 1923 BP.
 XX
 AC AAX37279;
 XX
 XX 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease encoding DNA.
 XX

KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 XX washing composition; oxidising agent; ss.
 OS Bacillus sp.
 XX WO9918218-A1.
 XX 15-APR-1999.
 PD 07-OCT-1998; 98WO-JP004528.
 PF 07-OCT-1997; 97JP-00274570.
 XX 07-OCT-1997;
 XX (KAO) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX P-PSDB; AAY17091.
 DR WPI; 1999-287736/27.
 DR P-PSDB; AAY17091.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX Disclosure; Page 63-68; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8,696-160 Length: 1923
 Score: 2120.50 Matches: 405
 Percent Similarity: 97.70% Conservative: 19
 Best Local Similarity: 93.32% Mismatches: 9
 Query Match: 94.16% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689a-7 (1-433) x AAX37279 (1-1923)
 QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPhcGlyLeuTyr 20
 Db 619 AATGATGTTGCGGTGAATGTTCAAGCGGATGTTGGCTCAGACGAGTACGGGTGAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAGGACAGATGATACCGGTTCGCGATACAGGCGTTCATACAGTTCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGTCATGAAGCTTCGCGGGAAATTAATGTCATTTATATGATTCGCGGACGAGT 798
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-- 79
 Db 799 AATGCCAATGATACGAAATGTTGATGCGCATGTTGGCTCCGTTATTAGGAACCGC 858
 QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 859 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCATCTATCATGATAGC 918
 QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119

Db 919 GGTGGGAGCTTGAGGACTACCTTGAATCTGCAACCTTATTACAGCAAGCATACAGT 978
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThr 139
 Db 979 GCTGTGCCAGATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGGCTTACACAA 1038
 QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCCCTTTTCGTGCC 1098
 QY 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 1099 GGGAAATGAAGACCGGAACCGCGGAACCATCAGTCCACCGACAGCAGTAAATATGCATA 1158
 QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
 Db 1159 ACAGTCGGAGCTACGGAAACCTTCGCGCAAGCTTTGGGTCTTATGCGGCAATATCAAC 1218
 QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
 Db 1219 CATGTGGCACAGTTCCTCTTCACGTGGACCGCAAGAGTGGACGGATCAACCGGATGTC 1278
 QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 1279 ATGGCACCGGAACGTTCTATCTACTATCAGCAAGATCTTCTTGCACCGGATTCCTCTTC 1338
 QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
 Db 1339 TGGCGGAACCATGACAGTAAATATGATACATGATGGTGGAGCTCCATGGCTACACCGATC 1398
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
 Db 1399 GTTGTGGAACGTTGGACAGCTTCGTGAGCATTTGTGAAAACAGAGGATCACACCA 1458
 QY 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
 Db 1459 AAGCTTCTCTATTAAAGCGCACTGATTCGCGTGCAGCTGACATCGGCTTGGCTAC 1518
 QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
 Db 1519 CCGAAGCGTAAACCAAGGATGGGAGCGATGACATTTGGATTAATCCCTGAAAGTTGCCTAT 1578
 QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339
 Db 1579 GTGAAGAGTCCAGTTCCTATCCACAGCCAAAGGAGCGTACTCGTTTACTGCTACT 1638
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359
 Db 1639 GCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTTCGAGGACCAACTGCT 1698
 QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379
 Db 1699 TCGGTAACTGTTGCAATGATCTGGACCTTGTTCATTACCGCTCCAAATGGCACAGTAT 1758
 QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
 Db 1759 GTAGGAATGACTTTTACTTCGCCATACAAATGATACTGGGATGGCGCGCAATTAACGTAG 1818
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
 Db 1819 AATGATTTTATTAATGCACCAACAAAGCGGAGTATACATTAAGTACAGGCTTATAAC 1878
 QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
 Db 1879 GTACCGGTGGACACAGAACTTCTCGTTGGCAATTTGGAAT 1920

RESULT 4
 AAX37277
 ID AAX37277 standard; DNA; 1920 BP.
 XX
 AC AAX37277;
 XX

DT 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease encoding DNA.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 07-OCT-1998; 98WO-JP004528.
 XX
 PR 07-OCT-1997; 97JP-00274570.
 XX
 PA (KAOS) KAO CORP.
 XX
 PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17087, AAY17089.
 XX
 PT Alkali protease from Bacillus used in washing powders.
 XX
 PS Disclosure; Page 53-58; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC Bacillus. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

 Alignment Scores:
 Pred. No.: 5,45e-159 Length: 1920
 Score: 2110.50 Matches: 404
 Percent Similarity: 97.24% Conservative: 18
 Best Local Similarity: 93.09% Mismatches: 11
 Query Match: 93.72% Indels: 1
 DB: 2 Gaps: 1

 US-09-985-689a-7 (1-433) x AAX37277 (1-1920)
 QY 1 AsnAspValAlaArgGlyLeValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
 DB 616 AATGATGTGCCAGAGATTGTCAAAAGCGGATGTGCCACAGCAGCTACCGTTTGTAT 675
 QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 676 GGACAGGCGAGATTGTCCGAGTTGCCGATCTGGATTGGATACAGAGAAACACGACGT 735
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 736 TCGATGATGAAGCCITTCGCGGTAAATAACAGCAGCTATATGCACTGGGTCCGAGCAAT 795
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValleuGlyAsn--- 79
 DB 796 AATGCGAATGATACGAACCGGTACCGGTACCGGTACCGGTACCGGTACCGGTACCGGTACCG 855

RESULT 5

QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 DB 856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGGTTTTTCAATCATCATGGATAGC 915
 QY 100 GlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
 DB 916 AGTGGTGGCTTGGAGGCTTGCCTTCCAACTCTCAAACTTATTCAGCAAGCATTCAGT 975
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139
 DB 976 GCAGGTGCCAGATTTCATCAAACTCTCTGGGGGCGAGCGGTGAATGGGGGCTACACACA 1035
 QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
 DB 1036 GATTCCAGAAATGTGGATGACTATGTAAGGAAATAATGATATGACGATCTTTTCGGGCT 1095
 QY 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 DB 1096 GGGAAATGAAGGCGGAAACGCGGTACCATCAGTCACCTGCTACGGCTAAACACGCCATA 1155
 QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
 DB 1156 ACAGTCGGCGCACACCGAAACCTCGCTCCAAAGCTTCGGTTCTCTATGAGATAATATTAAAC 1215
 QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
 DB 1216 CACGTTCACAGTTCTCTTCCCGTGGCCGACAAAGATGGCGAATCAAGCCTGATGTC 1275
 QY 220 MetAlaProGlyThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 DB 1276 ATGGCGCGAGGACATACATTTATCAGCAAGATCTTCTTTCGCCCGGATTCCTCCTTC 1335
 QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 DB 1336 TGGCGGATCATGACACCAATATGCTATATGGTGGACGTCCTCAAGCAACACCGATT 1395
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
 DB 1396 GTTTCGGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGAAATAGAGAAATCACTCCT 1455
 QY 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
 DB 1456 AAGCTTCCCTATTGAAGCAGCTTTGATTCAGAGTGCTGCTGATGTGGATTGGGTAT 1515
 QY 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
 DB 1516 CCGAACGAAACCAAGGATGGGCGCGAGTGACCTGGATAAATCGTTGAAACGTTGCCTAT 1575
 QY 320 ValAsnGluThrSerProLeuSerThrSerGlnIleAlaThrTyrSerPheThrAlaGln 339
 DB 1576 GTGAACGAATCCAGTGCCTTATCACTAGCCAAAGCAACGACATATACCTTTACTGCAACG 1635
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359
 DB 1636 CGGGGCAAGCCATTGAAATCTCCCTGGTATGGTGGATGCGCTCCCTGCAAGCACTACTGCT 1695
 QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379
 DB 1696 TCTGTAAACCTGGTCAATGATTTGGATTGGCTATTACAGCAACCAACGCAAGATAT 1755
 QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnTrpAspGlyArgAsnAsnValGlu 399
 DB 1756 GTCCGGATGACTTCTCAGCACCATTTGACATAACTGGGATGGCGCCCAATAACGTAGAA 1815
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
 DB 1816 AATGTATTATTAAATTCGCCCAAGTGGAAACATATATACCATTTAGGTAGGTCGAAGCATATAAT 1875
 QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
 DB 1876 GTCCGGTGGACCAACAAACTTCTCGTTGGCAATTTGTGAAC 1917

AAQ27516
 ID AAQ27516 standard; DNA; 1299 BP.
 AC
 XX AAQ27516;
 DT 05-FEB-1993 (first entry)
 XX
 XX Alkali-protease Ya enzyme gene.
 DE
 XX Alkali resistance; surface active agent resistance; detergency improver;
 KW ss.
 XX Bacillus sp. Y.
 OS
 XX Key Location/Qualifiers
 FH 1. 1299
 CDS /*tag= a
 FT
 FT
 FT
 XX JPO4197182-A.
 XX
 PD 16-JUL-1992.
 XX
 XX 28-NOV-1990; 90JP-00327110.
 XX
 XX 28-NOV-1990; 90JP-00327110.
 XX
 XX (LLOY) LION CORP.
 XX
 XX WPI; 1992-288440/35.
 DR P-PSDB; AAR26274.
 XX
 PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 PT resistance and improves detergency.
 XX
 XX Claim 3; Page 2; 17pp; Japanese.
 PS
 CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 CC in alkali resistance and surface active agent resistance and improves
 CC detergency
 XX
 XX Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6.2e-153 Length: 1299
 Score: 2032.00 Matches: 384
 Percent Similarity: 94.46% Conservative: 25
 Best Local Similarity: 88.68% Mismatches: 24
 Query Match: 90.23% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-7 (1-433) x AAQ27516 (1-1299)

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnAsnAspPheGlyLeuTyr 20
 Db 1 AATGATGACCAAGAGGATAGTAAGCTGATGTCACAAACAATACGGATTATAT 60
 QY 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAAAGTCACTAGTTGCAGTAGCGGACACAGGCTTAGATACAGGTCGTAACGATAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAAGCAATTCGCGGGAATATCAGAGCTTTACGCGTTAGGAGAACTPAT 180
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
 Db 181 AATGCGATGATCCGAATGGCATGGCACACATGTAGCAGGTTCTGTACTTGGTAATGCT 240
 QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
 Db 241 TTTAAATAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAAATCATATTATGATAGCAGC 300

RESULT 6

AAT85667

ID AAT85667 standard; DNA; 1977 BP.

XX

AC AAT85667;

QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
 Db 301 GGAGGATTAGGTGGCTTACCATCGAACTTAAATACGTTATTAGTCAAGCTTGAATGCT 360
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp 140
 Db 361 GGAGCAAGATTACATACTAATCTTTGGGAGCCCGAGTAAATGGAGCGTACACTGCTAAC 420
 QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
 Db 421 TCGAGACAGGAGGATGATGTTTCGAAATAATGATATGACGGTACCTTTTTCGAGCTGGT 480
 QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 481 AATGAGGTCCTAATTCAGGAACAATTAGTCTCCAGGTACAGCGAAAAATGCTATTACG 540
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
 Db 541 GTCGGCGCAACGGAAACTATCGCCCAAGCTTCGGTTCGATAGCAGATAAACCANAATCAT 600
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
 Db 601 AITGCACAAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCCTGACGTAACA 660
 QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 Db 661 GCTCCTGGAACATTTATTTATCAGCAGCTTCTTCTTAGCTCCAGACTCTTCGTTTGG 720
 QY 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 721 GCGAATTATACAGTAAATACGCTATATGGCGGTACTCTCCATGGCCACACCTATTGTT 780
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
 Db 781 GCAGGGAATGTCGCGCAATTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCTTAAG 840
 QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
 Db 841 CCTTCTTTAAATAAAGCTGCATTCGCTGGTGTACTGATGTTGGTTTAGGATATCCT 900
 QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
 Db 901 AGTGGTGACCAAGGCTGGGGGCGTGTACTCTAGATAAATCGTTAAATGTAGCGTATGTC 960
 QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340
 Db 961 AATGAAGCAACTGCATTAGCCACAGACAAAAAGCAAGCTATTCGTTTCAAGCACACGG 1020
 QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360
 Db 1021 GGTAAACCTTTAAAAATCTCGTTAGTAGACAGATGCTCCTGGAAGTACAACTGCATCT 1080
 QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380
 Db 1081 TATPACACTAGTTAATGATTAGATCTAGTTATCTAGTCTCCGAATGACAAAAATATGTA 1140
 QY 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
 Db 1141 GGAATGATTTAGTTATCTTATGATATAAATCGGATGGTGGCAACAATGTTGAGAAC 1200
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrThrValGluValGlnAlaTyrAsnVal 420
 Db 1201 GTATTTATAAACGCTCCGAATCTGGAACGTATATAATGAGGTTCAAGCGTATATGTA 1260
 QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
 Db 1261 CCATCTGGCCACAGCGTTCTCAGTATCGTACAT 1299

XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX Thermococcus protease coding sequence.
 DE Thermococcus protease coding sequence.
 XX Protease; research reagent; thermal stability; thermococcus celer; ss.
 XX Thermococcus celer; DSM-2476.
 OS WO9721823-A1.
 PN 19-JUN-1997.
 XX 07-NOV-1996; 96WO-JP003253.
 XX 12-DEC-1995; 95JP-00323285.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 PI Tsunashawa S, Kato I;
 XX WPI; 1997-332794/30.
 DR P-PSDB; AAW24121.
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX Claim 3; Page 86-87; 159pp; Japanese.
 CC This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 5.08e-25 Length: 1977
 Score: 432.50 Matches: 134
 Percent Similarity: 44.85% Conservative: 62
 Best Local Similarity: 30.66% Mismatches: 156
 Query Match: 19.21% Indels: 85
 DB: 2 Gaps: 16

US-09-985-689a-7 (1-433) x AAT85667 (1-1977)

QY 8 ValysAlaAspValAlaGlnAsnAsnPhGlyLeuTyrglyGlnGlyValAla 27
 Db 433 ATAGGGCCGATACCGTCTGGAATCCTCGGTACGAGGAGCGTGTGGTGGC 492
 QY 28 ValAlaAspThrglyLeuAspThrglyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCGTCGATACGGGTATAGCGCAAC-----CACCCCGATCTGAAG 534
 QY 48 GlyLysileThrAlaLeuTyri---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
 Db 535 GGCAGGTGATAGCTGTGACGCGCGTCAACGCGAGGTGCGACCCCTACGATGACCAG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82
 Db 595 GGACACGGAAACCCACGTTGGGGGTATCGTTGCCGGAACCGGACGCGTAACTCCCAAGTAC 654
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100
 Db 655 ATAGGGGTGCCCCCGCGCGAAGTCTCTCGGCTCAGGTTCTCGGTGCGGACGCTTCG 714
 QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyriSerAla 120

Db 715 GGAAGCGTCTCCACCATCATCGCGGGTGTGACTGGTCTGCAGAACAGGACAAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyriThrAsp 140
 Db 775 GGGATAAGGGTTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGAGCGAACCGACTCC 834
 QY 141 SerArgAsnValAspAspTyriValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
 Db 835 CTCAGTCAGGCGGTCAACAACGCTGGGACCGCGGTATAGTAGTCTCGTCCGCGCGGC 894
 QY 161 AsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 895 AACAGCGGCGGAACACCTACCGTGGTCTACCGCGCCCGCGAGCAAGGTGCATAACC 954
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyriAlaAspAsnIleAsnHis 200
 Db 955 GTCGGTGCA-----GTTGACACAAACGACAAC 981
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
 Db 982 ATCGCCAGCTTCTCCAGAGGGGACCGACCGGAGGCTCAAGCCGGAAGTCGTC 1041
 QY 221 AlaProGlyThrTyriIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 Db 1042 GCGCCGCGGTGACATCATAGCCCGCGCGCCAGC-----GGAACGACATCGGC 1092
 QY 241 AlaAsnHisAspSerLysTyriAlaTyriMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 1093 ACCCCGATAACGACTACTACCAAGCGCTCTGGACACGATGGCCACCCGACGTT 1152
 QY 261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
 Db 1153 TCGGCGGTGGCGGCTCATCTCCAGCGCCAC-----CCG 1188
 QY 280 Lys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
 Db 1189 AGCTGGACCCCGGAAGGTAAGACCGCCCTTCAGACCCCGGACATAGTCGCCCCC 1248
 QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgVal 309
 Db 1249 AAGGATAGCGGACATCGCTACGGTGG-----GGTAGGGTG 1287
 QY 310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326
 Db 1288 AACGTCTACAAGGCCATCAAGTACGACGACTACCGCAAGCTCACCTTCACCGGCTCCGTC 1347
 QY 327 SerThrSerGlnLysAlaThrTyriSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346
 Db 1348 GCGCAAGGGAAGCGCCACCCACCTTC----- 1377
 QY 347 SerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364
 Db 1378 -----GAGTCAGCGCGCCACCTTCGTGACCGCCACCCCTCTACTGGGAC 1422
 QY 365 -----AsnAspLeuAspValIleThrAlaProAsnGlyThrLysTyriValGly 381
 Db 1423 ACGGCTCGACGACATGACCTTACTCTACACCCCAACGGAAGAGAG---GTTGAC 1479
 QY 382 AsnAspPheThrAlaProTyriAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401
 Db 1480 TACTCTACACCGGCTACTAC-----GGCTTCGAGAAGGTC 1515
 QY 402 PheIleAsnAlaProGlnSerGlyThrTyriThrValGluValGlnAlaTyri 418
 Db 1516 GGCTACTACAACCGGACCGCGGAACTTGGACGCTCAAGTCTCAGCTAC 1566

RESULT 7

AA05926
 ID AA05926 standard; DNA; 1977 BP.
 XX
 AC AA05926;
 XX
 DT 06-MAY-1999 (first entry)

XX W09856926 Seq ID 11.
 XX DE
 XX KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX NW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX OS Thermococcus celer.
 XX PN W09856926-A1.
 XX PD 17-DEC-1998.
 XX PF 04-JUN-1998; 98WO-JP002465.
 XX PR 10-JUN-1997; 97JP-00151969.
 XX PA (TAKI) TAKARA SHUZO CO LTD.
 XX PI Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
 XX DR WPI: 1999-080907/07.
 XX DR P-PSDB; AAW94840.
 XX PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 XX PT gene encoding it, for large scale production of the protease for
 XX PT industrial use.
 XX PS Disclosure; Page 53-54; 82pp; Japanese.
 XX The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 90-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 5.08e-25 Length: 1977
 Score: 432.50 Matches: 134
 Percent Similarity: 44.85% Conservative: 62
 Best Local Similarity: 30.66% Mismatches: 156
 Query Match: 19.21% Indels: 85
 DB: 2 Gaps: 16

US-09-985-689A-7 (1-433) x AAX05926 (1-1977)

QY 8 VallysAlaAspValAlaGlnAsnAAsnPhGlyLeuTyGlyGlnIleValAla 27
 DB 433 ATAGGGCCGATACCGTCTGGAACTCCCTCGCTACGACGAGGCGGTGTGGTGGCC 492
 QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 DB 493 ATCTGTCGATACGGGTATAGACCGGAAC-----CACCCGATCTGAAG 534
 QY 48 GlyLysIleThrAlaLeuTyf---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
 DB 535 GGCNAGGTCATAGGCTGGTACGACCGCTCAACGCGAGGTGCGACCCCTACGATGACCG 594
 QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82
 DB 595 GGACACGGAAACCCAGTTTCGGGGTATCGTTGCCGGAACCGGACGAGCGTTAACTCCGAGPAC 654
 QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100
 DB 655 ATAGGGCTCGCCCGCGCGAAGCTCTGCGGCTCAAGGTTCTCGGTGCCGACGGTTCG 714

QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 120
 DB 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGGTCTGCAGACACAGGCAAGTAC 774
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAsp 140
 DB 775 GGGATAAGGGTCACTCAACCTCTCTCCCTCGCTCTCTCCAGAGCTCCGACGGAACGACTCC 834
 QY 141 SerArgAsnValAspAspTyfValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
 DB 835 CTCAGTCAGCGCTCAACACGCTGGAGCCCGGTATAGTAGTCTCGTCCGCCCGCC 894
 QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 DB 895 AACAGCGGGCGAACAACCTACACGCTCGGTCTACCCCGCGCGAGCAAGGTCTATAACC 954
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyfAlaAspAsnIleAsnHis 200
 DB 955 GTCGGTGCA-----GTTGACGCAACGACAAC 981
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
 DB 982 ATCGCCAGCTTCTCCAGCAGGGAGCCGACCGGAGGCTCAAGCCGGAAGTGGTC 1041
 QY 221 AlaProGlyThrTyfIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
 DB 1042 GCCCGCGCGTGTACATCATAGCCCGCGCGCCAGC-----GGMACCGATGGGC 1092
 QY 241 AlaAsnHisAspSerLysTyfAlaTyfMetGlyGlyThrSerMetAlaThrProIleVal 260
 DB 1093 ACCCGATAACGACTACTACACCAAGCCCTCGAACCGATGGCCACCCGACGCTT 1152
 QY 261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
 DB 1153 TCGGGCGTGTGGCGCTCATCTCCAGCCCGCCAC-----CCG 1189
 QY 280 Lys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 292
 DB 1189 AGCTGGACCCCGGACAGGTGAGACCCCTCATCGAGACCGCGACATAGTCGCCCC 1248
 QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTyfTrpGlyArgVal 309
 DB 1249 AAGGAGATAGCGGACATCGCTACCGTGGC-----GGTAGGGTG 1287
 QY 310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326
 DB 1288 AACGCTACAGGCCATCAAGTACGACGACTAGCGCAAGCTCACCTTACCGGCTCCGTC 1347
 QY 327 SerThrSerGlnLysAlaThrTyfSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346
 DB 1348 GCGACACAGGGAAGCGCCACCCACACCTTC----- 1377
 QY 347 SerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364
 DB 1378 -----GACGTACGCGCGCCACCTTCGTGACCCGCCACCTCTACTGGGAC 1422
 QY 365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyfValGly 381
 DB 1423 ACGGGCTCGAGGACATCGACCTTACCTCTACGACCCCAACGGGAACGAG---GTTGAC 1479
 QY 382 AsnAspPheThrAlaProTyfArgAsnAsnTrpAspGlyArgAsnAsnValGluAsnVal 401
 DB 1480 TACTCTACACCGCTACTAC-----GGCTTCGAGAGGTC 1515
 QY 402 PheIleAsnAlaProGlnSerGlyThrTyfThrValGluValGlnAlaTyf 418
 DB 1516 GGCTACTACACCCGACCGCGGAACCTGACGGTCAAGGTCTGCTACGCTAC 1566

RESULT 8
 AAX05920
 ID AAX05920 standard; DNA; 1236 BP.
 XX

AC AAX05920;
 XX
 DT 06-MAY-1999 (first entry)
 XX
 DE Hyperthermostable protease fragment encoding DNA.
 XX
 KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
 XX
 OS Pyrococcus furiosus.
 XX
 PN WO9856926-A1.
 XX
 PD 17-DEC-1998.
 XX
 PF 04-JUN-1998; 98WO-JP002465.
 XX
 PR 10-JUN-1997; 97JP-00151969.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 PI Takakura H, Morishita M, Shimajo T, Asada K, Kato I;
 XX
 DR WPI; 1999-080907/07.
 XX
 DR P-PSDB; AAW94836.
 XX
 PT Recombinant hyperthermostable protease from Pyrococcus furiosus - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.
 XX
 PS Claim 6; Page 37-38; 82pp; Japanese.
 XX
 CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
 CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIO-Ala-Gly-Gly-Asn-PRO, where SIO is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially Bacillus strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis
 XX
 SQ Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2.32e-22 Length: 1236
 Score: 396.00 Matches: 127
 Percent Similarity: 42.56% Conservative: 56
 Best Local Similarity: 29.53% Mismatches: 163
 Query Match: 17.58% Indels: 84
 DB: 2 Gaps: 15

US-09-985-689A-7 (1-433) x AAX05920 (1-1236)

QY 16 AsnPhcGlyLeuTyrglyGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
 DB 58 AACTGGGATGATGATGCTCTGGATCACAATAGGAATAATTGACATGGATGAC--- 114
 QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyra 55
 DB 115 -----GCTTCTCCAGATCTCCAGGAAAGTA----- 144
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 DB 145 ATTGGGGTGGGAGATTTTGGTCAATGGTAGGATTTATCCATGATGACCATGACATGGA 204
 QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
 DB 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGACGACCAAGTAATGCGAAGTACAGGGA 264

QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102
 DB 265 ATGGCTCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTCGCGATGTTCTTGGAGC 324
 QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrrSerAlaGlyAla 122
 DB 325 ATATCTATATAATTAAGGAGGATTTAGTGGGCCCTTGATTAACAAGATAAGTACGGGAAT 384
 QY 123 ArgIleHisThrAsnSerTrpGlyValaProValaAsnGlyAlaTyrrThrAspSerArg 142
 DB 385 AAGGTCAATTAATCTTTCTCTGGTTCAAGCAGAGCTCAGATGGTACTGACGCTCTAAGT 444
 QY 143 AsnValaAspAspTyrrValaArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
 DB 445 CAGGCTGTAATGCAGCGTGGGATGCTGGATTAGTTGTTGGTTGCGCTGGAAACAGT 504
 QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 DB 505 GGACCTAACAGTATACAATCGTTCTCCAGCAGCTCCAGCAAGAGTTTACAGTTGGA 564
 QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrrAlaAspAsnIleAsnHisValAla 202
 DB 565 GCC-----GTTGACAAAGTATGATGTTATAACA 591
 QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
 DB 592 AGCTTCTCAACAGCAGGCGCAACTGCAGCGCAGGCTTAAGCTGAGGTTGTTGCTCCA 651
 QY 223 GlyThrTyrrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 DB 652 GGAACTGGATAATGCTGCCAGAGCAAGT-----GGAACTAGCATGGGTCAACCA 702
 QY 243 HisAspSerLysTyrrAlaTyrrMetGlyThrSerMetAlaThrProIleValAlaGly 262
 DB 703 ATTAATGATGATTTACACAGCAGCTCCTGGGACATCAATGGCACTCTCAGTACGTGGT 762
 QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
 DB 763 ATTGAGCGCTCTTGTCTCAA-----GCACACCGAGCTGGACT 801
 QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
 DB 802 CCAGACAAGTAAACACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 861
 QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312
 DB 862 GCGGATATAGCCTACGGTGA-----GGTAGGGTTAATGCATAC 900
 QY 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329
 DB 901 AAGGCTATAAACTACGATAAATATGCAAAAGCTAGTGTCTACTGGATATGTGCCAACAAA 960
 QY 330 GlnLysAlaThrTyrrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349
 DB 961 GGCAGCCAACTCACCACTGCTGTTATAGCGGAGCTTCGTTAGCGAATGCCACATATAC 1020
 QY 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
 DB 1021 TGGGACATGGCAAT-----AGGACCTTGATCTT 1050
 QY 370 ValIleThrAlaProAsnGlyThrLysTyrrValGlyAsnAspPheThrAlaProTyrrAsp 389
 DB 1051 TACCTCTACGATCCCAATGGAAACACAG---GTTGACTACTCTTACACCGCTACTAT--- 1104
 QY 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
 DB 1105 -----GGATTCGAAAGGTTGGTTATTATACACCACTGATGGA 1143
 QY 410 ThrTyrrThrValGluValGlnAlaTyrrAsn 419
 DB 1144 ACATGGACAATTAAGGTTGTAAGTACAGC 1173

RESULT 9

AAT85668
 ID AAT85668 standard; DNA; 1566 BP.
 XX
 AC AAT85668;
 XX
 DT 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX
 DE Pyrococcus furiosus protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
 XX
 OS Pyrococcus furiosus; DSM-3638.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1566
 FT /tag= a
 FT /transl_except= (pos: 1282..1284, aa: Xaa)
 FT /note= "Xaa= Gly, Val"
 XX
 WO9721823-A1.
 XX
 PD 19-JUN-1997.
 XX
 PF 07-NOV-1996; 96WO-JP003253.
 XX
 PR 12-DEC-1995; 95JP-00323285.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 PI Tsunawasa S, Kato I;
 DR WPI; 1997-332794/30.
 DR P-PSDB; AAW24122.
 XX
 PT Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX
 PS Claim 7; Page 90-91; 159pp; Japanese.
 XX
 CC This sequence represents the coding sequence for the protease from
 CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX
 SQ Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;
 Alignment Scores:
 Pred. No.: 3,12e-22 Length: 1566
 Score: 396.00 Matches: 127
 Percent Similarity: 42.56% Conservative: 56
 Best Local Similarity: 29.53% Mismatches: 163
 Query Match: 17.58% Indels: 84
 DB: 2 Gaps: 15
 US-09-985-689A-7 (1-433) x AAT85668 (1-1566)
 QY 16 AsnPheGlyLeuTyrglyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
 Db 58 AACTTGGATATGATGTTCTGGATCAACATAGGAATTAATGACACTGGAATTGAC--- 114
 QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyValIleThrAlaLeuTyra 55
 Db 115 -----GCTTCTCATCCAGATCTCCAGGAAGAAGTA----- 144
 QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 145 ATTGGGTGGGTAGATTGTCATGTAGGAGTTATCCATGATGACCATGACATGACATGGA 204

QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
 Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAGGGA 264
 QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102
 Db 265 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAAGGTTCTAGTGCGGATGGTCTTGGAGC 324
 QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAla 122
 Db 325 ATATCTACTATAATTAAGGGAGTTGAGTGGCCGTTGATACAAAGATAAGTAAAGTAA 384
 QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSerArg 142
 Db 385 AAGGTCAATTAATCTTCTCTTGGTTCAAGCAGAGCTCAGATGGTACTGACGCTCTAAGT 444
 QY 143 AsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
 Db 445 CAGGCTGTTAATGACGCGTGGATCTCGATTAGTTGTTGGTTGCGCTGGAAACAGT 504
 QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 505 GGACCTAACACAGTATACATCGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGGA 564
 QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAla 202
 Db 565 GCC-----GTTGACAGTATGATGTTATAACA 591
 QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
 Db 592 AGCTTCTCAAGCAGAGGGCCAACTGCAGAGCGCAGGCTTAAGCCTGAGGTTGTTGCTCCA 651
 QY 223 GlyThrTyrlleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 652 GGAACACTGGATAATTGCTCCAGAGCAAGT-----GGAACTAGCATGGTCAACCA 702
 QY 243 HisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 703 ATTAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCACGTAGTGTGT 762
 QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
 Db 763 ATTGCAGCCCTCTTGTCTCAA-----GCACACCCGAGCTGGACT 801
 QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
 Db 802 CCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATCGTAAAGCCAGATGAAATA 861
 QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312
 Db 862 GCGGATATACCTTACGTTGCA-----GGTAGGTTAATGATATAC 900
 QY 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329
 Db 901 AAGCTATAAATACGATACATGCAAGCTAGTGTCTCAGTATGTTGCAACAAA 960
 QY 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349
 Db 961 GGCAGCCAAACTCACCAGTTCTGTTATTAGCGAGCTCGTTCGTAATCGCACATTATAC 1020
 QY 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
 Db 1021 TGGACAAATGCCAAT-----AGCGACCTTGATCTT 1050
 QY 370 ValIleThrAlaProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyThrAsp 389
 Db 1051 TACCTCTACGATCCCAATGGAACCCAG---GTTGACTACTCTTACACCGCTACTAT--- 1104
 QY 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
 Db 1105 -----GGATTCGAAAAGTGGTTATTACACACCACTGATGGA 1143
 QY 410 ThrTyThrValGluValGlnAlaTyAsn 419

RESULT 11

AA05929
ID AAX05929 standard; DNA; 1962 BP.
XX
AC AAX05929;
XX
XX
DT 06-MAY-1999 (first entry)
XX
XX
DE Hyperthermostable protease encoding DNA.
XX
XX
KW Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
KW additive; drug; washing agent; foodstuff; chemical synthesis; ds.
XX
OS Pyrococcus furiosus.
XX
XX
PN W09856926-Al.
PD 17-DEC-1998.
XX
XX 04-JUN-1998; 98MO-JP002465.
XX
XX 10-JUN-1997; 97JP-00151969.
XX
XX (TAKI) TAKARA SHUZO CO LTD.
XX
XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;
XX WPI; 1999-080907/07.
XX P-PSDB; AAW94841.
XX
XX Recombinant hyperthermostable protease from Pyrococcus furiosus - and
PT gene encoding it, for large scale production of the protease for
PT industrial use.
XX
XX Disclosure; Page 59-60; 82pp; Japanese.
PS
XX
CC The invention relates to a hyperthermostable protease derived from a
CC thermophilic bacterium (especially Pyrococcus furiosus). The protease has
CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at
CC 95 deg.C. The invention also provides gene sequences encoding a
CC polypeptide of formula SIG-Ala-Gly-Asn-PRO, where SIG is a signal
CC peptide from subtilisin, and PRO is the above protease. Host cells
CC (especially Bacillus strains) transformed with vectors comprising the
CC genes are used for the recombinant production of the protease. The
CC hyperthermostable protease which can be prepared in quantity suitable for
CC industrial use, can be used as an additive for drugs, washing agents and
CC foodstuffs and for chemical synthesis
XX
SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4,13e-22	Length:	1962
Score:	396.00	Matches:	127
Percent Similarity:	42.56%	Conservative:	56
Best Local Similarity:	29.53%	Mismatches:	163
Query Match:	17.58%	Indels:	84
DB:	2	Gaps:	15

US-09-985-689A-7 (1-433) x AAX05929 (1-1962)

QY	16	AsnPheGlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr	35
DB	454	AATTTGGGATATGATGGTCTGGATATCAATAGGATATATGACACTGGAAATGAC---	510
QY	36	GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla	55
DB	S11	-----GCTTCTCATCCAGATCTCCAGGAAAAGTA-----	540
QY	56	LeuGlyArgThrAsnAsnAlaen-----AspProAsnGlyHisGly	69
DB	541	ATTGGGTGGGTAGATTGTTGTCATGTGTAGGAGTTATCCATACGATGGACCATGGATGGA	600

QY	70	ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly	84
DB	601	ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTATGCAAGTACAGGGA	660
QY	85	MetalloProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly	102
DB	661	ATGGCTCAGGAGCTAAGCTGGCGGAATTAAAGCTTCTAGGTGCGGATGTTCTGGAGC	720
QY	103	LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla	122
DB	721	ATATCTACTATAATTAAGGGAGTTGAGTGGGCCCTGTATTAACAAGATAAGTAGCGGA	780
QY	123	ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg	142
DB	781	AAGGTCAATTAATCTTCTCTGGTTCAGCCAGAGCTCAGATGCTACTGACGCTCTAAGT	840
QY	143	AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu	162
DB	841	CAGGCTGTTAATGCAGCGTGGGATGCTGGATTAGTTGTTGTTGGTTCGCGTGGAAACAGT	900
QY	163	GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly	182
DB	901	GGACCTAAACAGTATACAAATCGTTCTCCAGAGCTGCAAGCAAAAGTTATTACAGTTGA	960
QY	183	AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla	202
DB	961	GCC-----GTTGCAAGTATGATTATTAACA	987
QY	203	GlnPheSerSerArgGlyProThrArgAspGlyValArgIleLysProAspValMetAlaPro	222
DB	988	AGCTTCTCAAGCAGAGGGCCAACTGCAGACGGCGCTTAAGCTGAGTTGTTGCTCCA	1047
QY	223	GlyThrTrpIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn	242
DB	1048	GGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACAGTATGATTATTAACA	1098
QY	243	HisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly	262
DB	1099	ATAATGACTATTACACAGCAGCTCTGGGACATCAATGGCACTCTCCAGTAGCTGGT	1158
QY	263	AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys-----	280
DB	1159	ATTGAGCCCTCTTCTCCAA-----GCACACCCGAGCTGGACT	1197
QY	281	ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla-----	292
DB	1198	CCAGCAAGTAAACACAGCCCTCATAGAACTCTCATATCGTAAAGCCAGATGAATA	1257
QY	293	AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp	312
DB	1258	GCCGATATAGCCTACGGTGCA-----GTTAGGTTAATGCATAC	1296
QY	313	LysSerLeu-----AsnValAlaPheValAsnGlnThrSerProLeuSerThrSer	329
DB	1297	AAGGCTATAAAGTACGATATATGCAAGCTATGTTCTAGTATGTTGCAACAAA	1356
QY	330	GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLysIleSerLeuVal	349
DB	1357	GGCAGCCAACTCACCACTGCTTATTAGCGAGCTTCGTTGCTGTAACACCATATATAC	1416
QY	350	TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu	369
DB	1417	TGGCAATATGCCAAT-----ACGACCTTATGCTT	1446
QY	370	ValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAsp	389
DB	1447	TACCTCTACGATCCCAATGGAAACAG-----GTTGACTACTCTTACACCGCTACTAT---	1500
QY	390	AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly	409
DB	1501	-----GGATTGAAAGGTTGTTATTATTACAACCCCACTGATGGA	1539

Qy 410 ThrTyThrValGluValGlnAlaTyrAsn 419
 Db 1540 ACATGGACAATTAAGGTTGTAGCTACAGC 1569

RESULT 12
 AAT85669
 ID AAT85669 standard; DNA; 1977 BP.
 XX AAT85669;
 AC
 XX
 DT 20-APR-1998 (first entry)
 XX
 DE Protease coding sequence.
 XX
 KW Protease; research reagent; thermal stability; ss.
 XX
 OS Synthetic.
 XX
 PN WO9721823-A1.
 XX
 PD 19-JUN-1997.
 XX
 PF 07-NOV-1996; 96WO-JP003253.
 XX
 PR 12-DEC-1995; 95JP-00323285.
 XX
 PA (TAKI) TAKARA SHUZO CO LTD.
 XX
 PI Takakura H, Morishita M, Yamamoto K, Mitra M, Asada K;
 PI Tsunaeawa S, Kato I;
 XX
 DR WPI; 1997-332794/30.
 XX
 P-PSDB; AAW24123.
 XX
 XX
 PT Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX
 XX Claim 11; Page 95-97; 159pp; Japanese.
 XX
 CC This sequence represents the coding sequence for a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries
 XX
 SQ Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,25e-21 Length: 1977
 Score: 390.00 Matches: 127
 Percent Similarity: 43.22% Conservative: 61
 Best Local Similarity: 29.20% Mismatches: 151
 Query Match: 17.32% Indels: 96
 DB: 2 Gaps: 17

US-09-985-689A-7 (1-433) x AAT85669 (1-1977)

Qy 16 AsnPheGlyLeuTyGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
 Db 454 AACTTGGGATATGATGTTCTTGGAAATCACAATAGGAATAATTGACACTGGAAATTGAC--- 510
 Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 511 -----GCTTCTCCAGATCTCCAGAAAGA--- 540
 Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGGTGGGTAGATTGTTCAATGGTAGGATTATCCATACGATGACCATGGACATGGA 600
 Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
 Db 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGCAAGTACACAGGA 660

RESULT 13

Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102
 Db 661 ATGGCTCAGAGCTAAGCTGGCGGAATTAAGTTCTAGTGCCGATGTTCTTGGAGC 720
 Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
 Db 721 ATATCTACTATAATTAAGGAGTTGAGTGGCCCTTGATAACAAAGATAAGTACGGAATT 780
 Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg 142
 Db 781 AAGGTCAATTAATCTTCTTGGTTCAAGCCAGAGCTCCGACGGAAACCGACTCCCTCAGT 840
 Qy 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
 Db 841 CAGCCCGTCAACAACGCTGGGACCGCGGTATAGTCTGCTGCGCGCGGCAACAGC 900
 Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGGCGGAACACCTACACCGCTCGGCTCACCCGCCCGGAGCAAGGTGTCATAACCGTGGT 960
 Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202
 Db 961 GCA-----GTTGACACCAACGACACATCGCC 987
 Qy 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
 Db 988 AGCTTCTCCAGCAGGGACCGACCGGACGGAAGGCTCAAGCCGGAAGTCTGCGCCCC 1047
 Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1048 GCGTTTGACATCATAGCCCGCGGCCAGC-----GGAACAGCATGGGCACCCCG 1098
 Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1099 ATAAACGACTACTACCAACAGCCTCTGGAACACGATGCGCACCCCGCAGCTTTCGGGC 1158
 Qy 263 ---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys--- 280
 Db 1159 GTTGGCGGCTCATCTCCAGGCCAC-----CCGAGCTGG 1194
 Qy 281 ---ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
 Db 1195 ACCCGGACAGGTGAGACCGCCCTCATCGAGCCGCGCATAGTCCGCCCAAGGAG 1254
 Qy 293 ---AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeu 311
 Db 1255 ATAGCGGACATCGCCTACGTCG-----GGTAGGTGAACGTC 1293
 Qy 312 AspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThr 328
 Db 1294 TACAAGGCGCATCAGTACGACGACTACGCCAAGCTCACCTTCACGGCTCGTCGCCGAC 1353
 Qy 329 SerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeu 348
 Db 1354 AAGGGAAGCGCCACCCACACCTTC----- 1377
 Qy 349 ValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364
 Db 1378 -----GAGTCAGCGGCCCGCCACTTCGTAGCCGCCACCTCTACTGGACACGGGC 1428
 Qy 365 ---AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAsp 383
 Db 1429 TCGAGCGACATCGACTCTACCTCTACGACCCCAACGGAACGAG-----GTTGACTACTCC 1485
 Qy 384 PheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIle 403
 Db 1486 TACACCGCCTACTAC-----GGCTTCGAAGAGTTCGGCTAC 1521
 Qy 404 AsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyr 418
 Db 1522 TACACCCGACCGCCGGAACCTGGAGCTCAAGGTCGTCAGCTAC 1566

```
ABL54900
ID ABL54900 standard; DNA; 2121 BP.
XX
AC ABL54900;
XX
DT 11-SEP-2003 (revised)
DT 31-MAY-2002 (first entry)
XX
DE T. yonsei subtilisin-like serine protease coding sequence.
XX
XX Subtilisin-like serine protease; ss.
XX
XX Thermoanaerobacter yonseiensis.
XX
XX Key Location/Qualifiers
XX CDS 142..1779
XX FT /*tag= a
XX FT /product= "subtilisin-like serine protease"
XX
XX KR2000072141-A.
XX
XX 05-DEC-2000.
XX
XX 04-AUG-2000; 2000KR-00045411.
XX
XX 04-AUG-2000; 2000KR-00045411.
XX
XX (KIMY/) KIM Y S.
XX
XX Chang HJ, Kim DH, Byun YR, Kim YS;
XX WPI; 2001-298092/31.
XX P-PSDB; ABB09483.
XX
XX New DNA sequence of thermophilic protein decomposition enzyme and protein
XX PT derived therefrom.
XX
XX Claim 1; Page 6; 15pp; Korean.
XX
XX This sequence represents the DNA encoding the Thermoanaerobacter yonsei
XX subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX to standardise OS field)
XX
XX Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;

Alignment Scores:
Pred. No.: 5,93e-20 Length: 2121
Score: 369.50 Matches: 118
Percent Similarity: 46.73% Conservative: 57
Best Local Similarity: 31.55% Mismatches: 136
Query Match: 16.41% Indels: 63
DB: 4 Gaps: 14

US-09-985-689A-7 (1-433) x ABL54900 (1-2121)
QY 6 GlyIleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTyrglyGlnGlyGlnile 25
DB 445 GGAATCACAAA-----GCAGAGTGTATTGGAGTCACAGGAAAAATATACA 495
QY 26 ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45
DB 496 ATAGCAATATTGACACAGGTATAGCGAAATACAGTTGACCTCTCA----- 543
QY 46 PheArgGlyIlylleThrAlaLeuTyrglyArgThrAsnAlaAsn----- 63
DB 544 ---GGTGGAAAAATA-----ATAGATGGAAGAGCTTTATCATCAACACAAA 585
QY 64 -----AspProAsnGlyHisGlyThrHisVal-----AlaGlySer 75
DB 586 ACTACACCATACGACCAATGCCATGGCACTCAGTAGCAAGTATTGCTCGAGGTACA 645
QY 76 ValLeuGlyAsnAlaThrAsnIysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 95
DB 766 GTTCAGGTTATCTGCCC---GGTTCGTACAGTGATCTGGAGCTTTTAACGCAACA 1416

646 GGTGCTGGAACAGTCTTTACAAAGCGTTGCTCTGATGCTTTTGTGGTAGAATAAAA 705
96 IleMetAspSer-----GlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeu 113
706 GTTTTAGATCAAAATGGAAGCGCAGCATGAGCAGTGTACTGACGAANTGACTGGCT 765
114 PheSerGlnAlaTyrsrAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProVal 133
766 GTTCAAAATAAGATGTATACGGAATCAAAAGTTATAAATTTAAAGCCTCGGCACTTCACA 825
134 AsnGlyAlaTyrrThrAspSerArgAsnValAspAspTyrrValArgLysAsnAspMet 153
826 AGTTCTGATGGAAGTCTCTACCTCATTCAGCAGTGTATAGACAGTAGATAGCGGTATT 885
154 ThrIleLeuPheAlaAlaGlyAsnGluGlyProGlySerGlyThrIleSerAlaProGly 173
886 GTAGTAGTTGTGACGAGCAAGAACTCTGGCCCTGCAAAATACACCATAGGTCCTCCCTGGT 945
174 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 193
946 GCTGCGGAAAAAGCCATAACACAGTCGACCAATGCGCAGATGTA-----GGTGAA 993
194 TyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrArgAspGly 213
994 CTGGCTTTAACCTT-----GCAAGCTTTCCAGCGCGTCTCTACTGCTGACGGA 1044
214 ArgIleLysProAspValMetAlaProGlyThrTyrrIleLeuSerAlaArgSerSerLeu 233
1045 AGATAAAAAACCTGACATTGCGGCCCCAGGATATATAATTAAGTCCGCGAAG----- 1095
234 AlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrrAlaTyrrMetGlyGlyThr 253
1096 -----GCGAATTCTGTAATGGATATGTAACATACACGCGGTACA 1134
254 SerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLys 273
1135 AGCATGGCAACACCTTTTGTAGCAGAACTGTGTCTTATGCTTAAC-----GCT 1185
274 AsnArgGlyValThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAla 293
1186 AATCCAAATCTCCTCCAAATGATGCA-----AAAAATATAATTAATGCTACTGCAAAA 1239
294 AspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLys 313
1240 AGCTGGGGCCCTCCAAGCAAAACGTTGACTACTGTGCGAGGAAGATTGGATGGCTATGAA 1299
314 SerLeuAsnValAla-----Phe 319
1300 GCTATAAGGGTAGCAGGTAAATTTAGAGGAATAATATTGATGTACCAATCATTTATTAT 1359
320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrrSerPheThrAla--- 338
1360 ATTTCCAGTTATCTGCCC---GGTTCGTACAGTGATCTGGAGCTTTTAACGCAACA 1416
339 GlnAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp 352
1417 AATACAAGTTATCCAAATCGCAATAACATTAATAATTCCTGAC 1458

RESULT 14
AAT61454
ID AAT61454 standard; DNA; 2539 BP.
XX
AC AAT61454;
XX
XX 06-OCT-1997 (first entry)
XX
XX Streptomyces viridosporus dhpA gene.
XX
XX asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
XX derivative; Streptomyces viridosporus; ester; chiral; synthesis;
XX cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
XX
XX Streptomyces viridosporus.
XX OS
```


XX AAT61455;
 AC 17-OCT-2003 (revised)
 DT 06-OCT-1997 (first entry)
 XX DhpA-mel chimeric gene.
 DE asymmetric hydrolase; dhpA; 4-substituted-1,4-dihydropyridine;
 KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
 KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
 XX Streptomyces viridosporus.
 OS Streptomyces antibioticus.
 OS Chimeric.
 XX Key Location/Qualifiers
 FH 338..2809
 FT CDS /tag= a
 FT mat_protein 338..2539
 FT /tag= b
 FT /product= "dhpA protein product"
 FT /note= "from S. viridosporus"
 FT 2540..2809
 FT /tag= c
 FT /product= "Melanin"
 FT /note= "from S. antibioticus"
 XX WO9705243-A1.
 XX 13-FEB-1997.
 XX 30-JUL-1996; 96WO-JP002147.
 XX 31-JUL-1995; 95JP-00212975.
 XX 29-FEB-1996; 96JP-00067478.
 XX (SAOC) MERCIAN CORP.
 XX Arisawa A, Matsufuji M, Teuruta T, Dobashi K, Nakashima T;
 PI Isshiki K, Yoshioka T;
 XX WPI; 1997-145682/13.
 DR P-PSDB; AAM13668.
 XX Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
 PT on 4-substituted-1,4-dihydropyridine derivatives to produce chiral
 PT derivatives useful for synthesis of cardiovascular drugs.
 XX Claim 5; Page 37-43; 78pp; Japanese.
 CC This sequence is a fusion gene encoding Streptomyces viridosporus dhpA
 CC gene product (an asymmetric hydrolase) which acts on 4-substituted-1,4-
 CC dihydropyridine derivatives, and melanin from S. antibioticus. The dhpA
 CC enzyme allows the efficient conversion of 4-substituted-1,4-
 CC dihydropyridine esters to chiral partially hydrolysed derivatives, for
 CC use in the synthesis of cardiovascular drugs suitable for the treatment
 CC of e.g. hypertension and ischaemic heart disease. (Updated on 17-OCT-2003
 CC to standardise OS field)
 XX SQ Sequence 2809 BP; 473 A; 1041 C; 984 G; 311 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1,89e-16 Length: 2809
 Score: 327.50 Matches: 136
 Percent Similarity: 40.83% Conservative: 51
 Best Local Similarity: 29.69% Mismatches: 160
 Query Match: 14.54% Indels: 111
 DB: 2 Gaps: 23
 US-09-985-689A-7 (1-433) x AAT61455 (1-2809)
 QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArg 37

Db 998 GGCTACGACGGCAAGGGGTGAAGATCGCGTCTGGACACCGGTGTGACACGAGC--- 1054
 QY 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
 Db 1055 -----CATCCGACCTGAAGGGCGGGTGCACCGCGCTCAAGAACTTCACC 1099
 QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
 Db 1100 GCGCGCGCGCGCGCGGACAAAGTGGCCACGACCGACCGCTCGATCGCGCGG 1159
 QY 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
 Db 1160 GGCACGGCGCGCGCGCTCCCAAGGGCAAGTACAAAGGGCGTTCGACCGCGCGCGATCCTC 1219
 QY 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110
 Db 1220 AACGGCAAGGTCTCTCGACGACTCCGGTTTCGGGACGACTCCGGCATCTCGCGCGCATG 1279
 QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
 Db 1280 GAGTGGCGCGCGCGCAG-----GGCCCGACGCTGCTCAACATGAGCTCGGC 1327
 QY 131 Ala-----ProValAsnGlyAlaTyrThrThrAspSerArgAsn 143
 Db 1328 GGCATGGACACACCGGACCGACCGCTGGAGCGCGG----- 1366
 QY 144 ValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGly 163
 Db 1367 GTCGACAAAGCTGTCGCGGAGAGGGGCTCTGTCGCTCGCGCGCGCGGCAAGAGGCG 1426
 QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
 Db 1427 CCGGAGTGG-----ATCGGTTCGCGGACGCGGACGCGCGCTCAGCGGAGCTCAGCGTCC 1480
 QY 184 ThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 203
 Db 1481 -----GTCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1507
 QY 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMetAlaPro 222
 Db 1508 TTCTCTCTCACCGCGCGCGCGCTCGCGGACGCGCGCTCAAGCGGAGCTCAGCGTCC 1567
 QY 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1569 GCGGTGGACATCAGCGCGCTCGCGGAGGGCAACGACATCGCGCAGGAGGTGGTGAG 1627
 QY 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1628 GACCGCGCGCGCTACATGACCATCTCCGCGCAGCTCGATGGCAGCCCGCGCAGCTCGCGGC 1687
 QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282
 Db 1688 GCGCGCGCGCTCTCGACGACGAG-----CACCCCGAC 1720
 QY 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
 Db 1721 TGGACCTCCGCGAACTGAAGGGCGCGCTCACCAGGCTCCACCAAGGGGCGCAATACACC 1780
 QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
 Db 1781 CCG---TTCGACGAGGTTCGGCGCGATCCAGCGCGACAAAGCGCTCCAGCAGACCGTG 1837
 QY 320 ValAsnGluThrSerProLeuSerThrSer----- 329
 Db 1838 ATC-----GCGGACCGCGTCTCGGTGAGCTTCGCGGTCCAGCAGTGGCGCACACCGAC 1891
 QY 330 -----GlnLysAlaThrThrSerPheThrAlaGlnAlaGlyLysProLeu 344
 Db 1892 GACGAGCGGTCCACGACGAGCTGACCTACCGAACCTTCGCGACCTCGGCACCGAGGAGCTG 1951
 QY 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354

Mon Apr 5 16:03:43 2004

```
Db 1952 AGCTGACGTGACCGCCACCGACCCCAAGGCAAGCGCGCCCGGGCTTCTTCAGG 2011
Qy 355 ---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 2012 CTGGGGCCACACG-----GTGACCGTC 2035
Qy 374 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 393
Db 2036 CCGCGCGCGCGCAGC--GCCTCCGTCGACATGACCGCC-----GACACCGCGCTCGGC 2086
Qy 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412
Db 2087 GGCACGTCGACCGCGGTACTCGCGGTACGTGTCGCC--ACGGCGCGCGCGCAGACG 2143
Qy 413 Val-----GluValGlnAlaTyrAsnValProVal 422
Db 2144 GTCCGACCGCCCGCGGTGACGCGAGGTGAGTCGTACGACGTGACCGTC 2197
```

Search completed: April 4, 2004, 01:13:19
Job time : 363.908 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 16:47:19 ; Search time 3544.27 Seconds
(without alignments)
5295.174 Million cell updates/sec

Title: US-09-985-689A-7

Perfect score: 2252

Sequence: 1 NDVARGIVKADVAQNFGLY.....EVQAYNPVSPQTFSLAIVH 433

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_epool/US09985689/runat_31032004_161806_4155/app_query.fasta_1.3498
-DB=GenEmbl -QWTF=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09985689 -CGCN_1_1_1925_@runat_31032004_161806_4155 -NCPU=6 -ICPU=3
-NO_WMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.ml.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2252	100.0	3003	6	AR059954	Sequence
2	2252	100.0	3003	6	BD062155	Nucleic a
3	2130.5	94.6	1923	1	AB084155	Bacillus
4	2125.5	94.4	1305	6	AX839476	Sequence
5	2125.5	94.4	1923	1	AB051423	Bacillus
6	2125.5	94.4	1923	6	AR368117	Sequence
7	2120.5	94.2	1923	6	AR368118	Sequence
8	2116.5	94.0	1302	1	AB046406	Bacillus
9	2110.5	93.7	1920	1	AB046403	Bacillus
10	2110.5	93.7	1920	6	AR368116	Sequence
11	2044	90.8	1299	1	AB046405	Bacillus
12	2044	90.6	1299	1	AB046402	Bacillus
13	2033	90.3	1299	1	AB046404	Bacillus
14	2032	90.2	2218	6	E03808	Sequence
15	626.5	27.8	60005	1	AF268611	Unculture
16	568.5	25.2	301205	1	AE017218	Gobacter
17	501	22.2	5772	3	AF466309	Dictyoste
18	499	22.2	5890	3	DDU60086	Dictyostell
19	494.5	22.0	6115	3	DDU20432	Dictyostell
20	445.5	19.8	5109	3	AF263455	Dictyoste
21	432.5	19.2	1977	6	AR201152	Sequence
22	423	18.8	126928	2	AC096673	Trypanoso
23	404	17.9	14376	1	AE013049	Thermoana
24	396	17.6	1235	6	AR201146	Sequence
25	396	17.6	1235	6	AR201155	Sequence
26	396	17.6	12452	1	AE010265	Pyrococcu
27	377	16.7	299850	1	AP004601	Oceanobac
28	374.5	16.6	303450	1	SC0939130	Streptomy
29	355	15.8	300425	1	AP005044	Streptomy
30	341.5	15.2	3348	1	D83672	Streptomyce
31	339.5	15.1	1239	1	AF305633	Thermoana
32	339.5	15.1	1239	1	AY028704	Thermoana
33	339.5	15.1	12295	1	AE013026	Thermoana
34	334.5	14.9	292550	1	AP001513	Bacillus
35	332.5	14.8	2539	6	AR202322	Sequence
36	332	14.7	302300	1	AP005034	Streptomy
37	329.5	14.6	291000	1	SC0939105	Streptomy
38	327.5	14.5	2809	6	AR202321	Sequence
39	327.5	14.5	3900	1	AB007809	Streptomy
40	326.5	14.5	300800	1	SC0939112	Streptomy
41	325	14.4	1329	6	AX433519	Sequence
42	323.5	14.4	207829	1	BSU80010	Bacillus su
43	323	14.3	135638	1	AF484556	Streptomy
44	322	14.3	6854	1	BSU99230	Bacillus sp
45	320	14.2	2731	1	AF015225	Thermococ

ALIGNMENTS

RESULT 1

```
AR069954
LOCUS       AR069954
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION   AR069954
VERSION     AR069954.1
KEYWORDS    1 GI:7220842
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 3003)
AUTHORS     Sloma,A. and Christianson,L.
TITLE       Nucleic acids encoding a polypeptide having protease activity
JOURNAL     Patent: US 5891701-A 41 06-APR-1999;
            Location/Qualifiers
FEATURES             source
                     1..3003
                     /organism="unknown"
                     /mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.:      2,23e-133      Length:      3003
Score:          2252.00        Matches:      433
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:     100.00%      Indels:         0
DB:              Gaps:         0

US-09-985-689A-7 (1-433) x AR069954 (1-3003)
QY      1 AsnAspValAlaAArgGlyVleVallyAlaAAspValAlaGlnAsnAsnPhGlyLeuTyr 20
DB      1470 ANTACGCTGGCCGTGGCATTTGTGAAGCAGACGCTGCACAAAATACTTTGGCTTATAT 1529
QY      21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB      1530 GGCAAGACAGACAGATTGTAGCAGTTGCTGATATCGGCTTGATACAGAAAGATGACGT 1589
QY      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB      1590 TCGATGCAATGAGCATTCGCGGTAGATTACCGCACATATATGCTGCGGCAGACGAAT 1649
QY      61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
DB      1650 AACGCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTTAGGAATGCT 1709
QY      81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
DB      1710 ACAATAAAGGATGGCACCAGCCCAATCTAGCTTTCAATCTATATGATGATGCT 1769
QY      101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
DB      1770 GGAGGGCTGGGAGGACTACCTGCTAATCTACAAACATTTATTCAGTCAAGCATATAGTCT 1829
QY      121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrAsp 140
DB      1830 GGAGCGAATTCATACCAATTCATGGGGGGCTCCAGTAAACGGTGCCTATACACAGAC 1889
QY      141 SerA:GAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
DB      1890 TCTCGAATGTTGATGATTATGTGAGAAAATAATGATATGACGATCTTTTTCGGCGCGGA 1949
QY      161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
DB      1950 AATGAGGGACCAAGTGGCGGTCAATTCAGTCAGCCAGGAAACAGCAAAATAATGCGATTACA 2009
QY      181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
DB      2010 GTTGGGGCAACCGAAACCTAGCTCCAGCTTCGATCTTATGCGGATTAATTAACCAT 2069
QY      201 ValAlaGlnPheSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
DB      2070 GTTGTCAATCTCTTCACGAGGTCCTACTAGAGATGGACGATTATTAAGCGGACGTCATG 2129
```

```
QY      221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
DB      2130 GCACACAGTACGATATATCTCTCTGTAGATCATCATTAGCTCCAGATTCCTATCTCTGG 2189
QY      241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
DB      2190 GCACACCATGATAGTAATATGCCTACATGGTGGTACTTCTATGGCTACTCCAAATTGTA 2249
QY      261 AlaGlyAsnValAlaGlnLeuArgGluHisPheVallyAsnArgGlyValThrProLys 280
DB      2250 GCAGGTAATGTGTGACCAATTAAGGGAGCATTTTGTGAAATAATAGAGGGTAACCTCTAAG 2309
QY      281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
DB      2310 CTTTCCCTTTTAAAGCTGCTTTAATTGCAGTCTCTGGGATGTTGGACTTGGCTTTCCA 2369
QY      301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
DB      2370 AATGTTAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTCGCAATTTGTG 2429
QY      321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340
DB      2430 AATGAACAGACGCCCTTTATCAACAGTCACAAAGCAACATATTCGTTACGGCTCAAGCT 2489
QY      341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360
DB      2490 GGTAAACCTTTAAATAATATCACTTTGTTGGTCAGATGCACCGGTAGCACCGCATCA 2549
QY      361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380
DB      2550 CTAACTTTAGTGAATGATTTAGATTAGATTATCTACCTGCACCAATGGAACTAAATACGTC 2609
QY      381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
DB      2610 GGAAATGACTTTTACAGCACCGGTATGATAAACAATGGGATGGCAGAAACAACGTTGAAAAT 2669
QY      401 ValPheIleAsnAlaProGlnSerGlyThrThrValGluValGlnAlaTyrAsnVal 420
DB      2670 GTGTTTATCAATGCTCTCCAAAGCGGAACGATATACGTCGAAGTCGAGCTTACATGTA 2729
QY      421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
DB      2730 CCAGTAAGTCCGCAACCTTTTCTTTAGCATTTGTACAT 2768

RESULT 2
BD062155      3003 bp      DNA      linear      PAT 27-AUG-2002
LOCUS         Nucleic acids encoding a polypeptide having protease activity.
DEFINITION
ACCESSION     BD062155
VERSION       BD062155.1 GI:22607760
KEYWORDS      JP 2001514529-A/39.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE     1 (bases 1 to 3003)
AUTHORS       Sloma,A. and Christianson,L.
TITLE         Nucleic acids encoding a polypeptide having protease activity
JOURNAL       Patent: JP 2001514529-A 39 11-SEP-2001;
            NOVO NORDISK BIOTECH INC
            PN JP 2001514529-A/39
COMMENT       PD 11-SEP-2001
            PF 09-JUN-1998 JP 1999503145
            PR 12-JUN-1997 US 08/873479
            PI ALAN SLOMA,LYNNE CHRISTIANSON
            PC C12N15/57,C12N15/75,C12N9/54,C12K14/00
            CC Strandedness: Single;
            CC Topology: Linear;
            PH Key Location/Qualifiers.
            Location/Qualifiers
            1..3003
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
```

FEATURES
source

1

ORIGIN

Alignment Scores:

Alignment Scores:	6.3e-126
Pred. NO.:	2130.50
Score:	97.93%
Percent Similarity:	93.78%
Best Local Similarity:	94.60%
Query Match:	1
DB:	1

US-09-985-689A-7 (1-433) x AB084155 (1-1923)

Qy	1	AsnAspValala	hrgGlylleVallyAlaAspValAlaGlnAsnAspPheGlyLeu	20
Db	619	AATGATGTTGCACGTCGGAATTTG	CAAAGCGGATGTGGCGCAGCAGCATACGGTGTGAT	678
Qy	21	GlyGlnGlyGlnIleValaValaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer	40	
Db	679	GGACAGGACAGATCGTAGCGGTTCGGATACAGGCTTGATACAGGTTCGCAATGACAGT	738	
Qy	41	SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaIleGlyArgThrAsn	60	
Db	739	TCGATGCATGAAGCCTTCGGGGGAAAATTA	CTGCATTATATGCAATGGGACGGACGAAT	798
Qy	61	AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValleuGlyAsn---	79	
Db	799	AAITGCCAATGATACGAATGTCATGTACGCATGTGCTGCTCCGTCCGTATTAGAAACGGC	858	
Qy	80	AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer	99	
Db	859	TCCACTAATAAAGGAATGGCGCTC	TACGGCGAACTCTAGTCTTCCAATCTCATGTGATAGC	918
Qy	100	GlyClyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySer	119	
Db	919	GGTGGGGGACATTGGAGGACTACCTT	CGAACTCGCAACCTTATTACGCCAAGCATACAGT	978
Qy	120	AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyThrThr	139	
Db	979	GCTGGTGCAGAAATTCATAAACTCCTCGGGGACGACAGTGAATGGGCTTACACAACA	1038	
Qy	140	AspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAla	159	
Db	1039	GATTCACGAATGTGGATGACTATGTGGCGAATAATGATATGACGATCCTTTTCGGTGCC	1098	
Qy	160	GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle	179	
Db	1099	GGGAATGAAGGACCGAAGCGGGAACCATCAGTGCACAGGCACAGCTMAAATGCAATA	1158	
Qy	180	ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn	199	
Db	1159	ACAGTCGAGCTACGGAAAACTCGCCCAAGCTTCGGGTCTTATGCGGACAATATCAAC	1218	
Qy	200	HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal	219	
Db	1219	CATGTGGCACAGTCTCTTCACGTGGACCGACAAAGGATGACGATCAAAACCGGATGC	1278	
Qy	220	MetAlaProGlyThrTyrlleuSerAlaArgSerSerIleuAlaProAspSerSerPhe	239	
Db	1279	ATGGCACCGGGAACCTTCATACTACACAAGATCTTCTCTGTGACCGGATTCCTCCTTC	1338	
Qy	240	TyrAlaAsnHisAspSerLysTyThrAlaTyThrMetGlyThrSerMetAlaThrProIle	259	
Db	1339	TGGGCGAACCATGACAGATAAATATGCATACATGGTGGAAACGTCCATGGGTACACCGATC	1398	
Qy	260	ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro	279	
Db	1399	GTTGTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA	1458	
Qy	280	LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe	299	
Db	1459	ALGCTTCTCTATTAAAGCGCACTGATTGCGGTGCGACAGACATCGGCTTGGCTAC	1518	
Qy	300	ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe	319	

Db	1519	CCGAACGGTAAACCAAGATGGGACGAGTGCACATTGGATTAATCCCTAAACGTTGCCTAT	1578
Qy	320	ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln	339
Db	1579	GTCAACGAGTCCAGTCTCTATCCACGACGCAAAAAGCAGCTACTCGTTTACTGTGCTACT	1638
Qy	340	AlaGlyValProLeuLysIleSerLeuValTyrSerAspAlaProGlySerThrThrAla	359
Db	1639	GCCGGCAAGCCTTTGAAGATCTCCCTGGTATGTCTTGATGCCCTCGCGACCAACAATGCT	1698
Qy	360	SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr	379
Db	1699	TCGGTAAACGCTTGTCAATGACCTTGGACCTTGTTCATTACCGCTCCAAATGSCACACAATAT	1758
Qy	380	ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyVargAsnAsnValGlu	399
Db	1759	GTTTGGAAATGACATTACTTCGCCATACAAATAATTAACCTGGGATGGCGCAATACGTAGAA	1818
Qy	400	AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn	419
Db	1819	AAAGTATTATTAAATGGCCCAACAGCGGACGCTATACAATTGAGGTACAGGCTTATAAC	1878
Qy	420	ValProValSerProGlnThrPheSerLeuAlaIleValHis	433
Db	1879	GTACCGGTTGGACACACAGACCTTCTCGTTGGCAATTGTGAAC	1920

RESULT 4

AX839476	AX839475	1305 bp	DNA	linear	PAT 15-DEC-2003
LOCUS	Sequence 2 from Patent EP1347044.				
DEFINITION	AX839476				
ACCESSION	AX839476.1	GI:39922766			
VERSION					
KEYWORDS					
SOURCE	Bacillus sp. KSM-KP43				
ORGANISM	Bacillus sp. KSM-KP43				
REFERENCE	Bacteria, Firmicutes; Bacilliales; Bacillaceae; Bacillus.				
AUTHORS	Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K., Saeki, K.K., Kobayashi, T.K. and Nomura, M.K.				
TITLE	Alkaline protease				
JOURNAL	Patent: EP 1347044-A 2 24-SEP-2003;				
FEATURES	Kao Corporation (JP)				
	Location/Qualifiers				

ORIGIN

Alignment Scores:

Pred.	No.: 8	Length:	1305
Alignment Scores:		Matches:	406
Score:	215.50	Conservative:	19
Percent Similarity:	97.93%	Mismatches:	8
Best Local Similarity:	93.55%	Indels:	1
Query Match:	94.38%	Gaps:	1
DG:	6		

UUS-09-985-689A-7 (1-433) x AX839476 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuValVallyAlaAspValAlaGlnAsnAsnPhGlyLeuTyr 20
 Db 1 AATGATGTTGGCGGTGGAATTTGTCAAAGCGGATGTGGCTCAGACGAGCTACGGGTTGTAT 60
 QY 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
 Db 61 GGCACAGGACAGATCGTAGCGGTTGCCGATACAGGGCTTGATACAGGTCGCAATGACAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyValThrAsn 60
 Db 121 TCGATGATGAAGCCCTCCCGCGGAAATTAATGCAATATATATGCAATGGGACGACCAAT 180
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 Db 181 AATGCCAATGATACGATGCTCATGTGACGATGTGGCTGCTCGTATAGGAAACGGC 240
 QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 Db 241 TCCACTAATAAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAATCTATCATGGATAGC 300
 QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
 Db 301 GTGGGGGACTGGAGGACTACTCTCGAATCTGCAAACTTATTCAGCCAGCAATACAGT 360
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThr 139
 Db 361 GCTGTGTCAGAAATTCATAAACTCCTGGGAGCAGCAGTCAATGGGGCTTACACACA 420
 QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
 Db 421 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGAGATCCCTTTCGCTGCC 480
 QY 160 GlyAsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 Db 481 GGGAAATGAAGGACCGAAGCGGCAACCATCATGTCACCGAGCAGCAGTCAAAAATGCAATA 540
 QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
 Db 541 ACAGTCGAGACTACGGAAACCTCCGCCCAAGCTTGGCTCTTATGCGGACATATCAAC 600
 QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyAglileLysProAspVal 219
 Db 601 CATGTGGCACAGTCTCTTTCAGTGCACGACGCAAGAGATGAGCGGATCAACCGGATGTC 660
 QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
 Db 661 ATGGCACCGGAAAGTTCATCTATCATCATCATCATCATCATCATCATCATCATCATCTTC 720
 QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyLysThrSerMetAlaThrProIle 259
 Db 721 TGGGGCAACCATGACAGATAAATATGATACATATGATGATGATGATGATGATGATGATGAT 780
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
 Db 781 GTTGTGGAAGAGTGGCACAGTCTGTGAGCATTTTGTGAAACACAGAGGATCACACCA 840
 QY 280 LysProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
 Db 841 AAGCCTCTCTATTAAGCGGCACTGATGCGGTCGAGCTGATGATGATGATGATGATGATGAT 900
 QY 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
 Db 901 CCGAAGCGTAACAGGATGGGAGGAGTGCATTTGGATTAATCCCTGGAAGCTTGCCCTAT 960
 QY 320 ValAsnGlyThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339
 Db 961 GTGAACGAGTCCAGTCTCTATCCACCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359
 Db 1021 GCGGCAAGCCCTTGAATATCTCCCTGGTATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 1080

QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProGlnGlyThrLysTyr 379
 Db 1081 TCCGTAAACGCTTGTCAATGATCTGACCTTGTCTATACCGCTCCAAATGGCACACAGTAT 1140
 QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
 Db 1141 CTAGGAATGACTTACTTCGCCATACAAATGATAACTGGGATGGCCGCAATACGTAGAA 1200
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
 Db 1201 AATGTATTATTATATGACCAAAAGCGGACGCTATACAAATGAGGTACAGGCTTAAAC 1260
 QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
 Db 1261 GTACCGGTGGACACAGACCTTCTCGTTGGCAATGTGAAT 1302

RESULT 5
 AB051423 1923 bp DNA linear BCT 10-MAY-2002
 LOCUS
 DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.
 ACCESSION AB051423
 VERSION AB051423.2 GI:20521154
 KEYWORDS
 SOURCE Bacillus sp. KSM-KP43
 ORGANISM Bacillus sp. KSM-KP43
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 Itoh, S. and Saeki, K.
 AUTHORS new protease
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 1923)
 REFERENCE
 AUTHORS Saeki, K.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-2000) Katsuhisa Saeki, KAO, CORPORATION;
 2606, AKABANE, ICHIKAIMACHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 3871185@kaxanet.kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 COMMENT On May 9, 2002 this sequence version replaced gi:14164344.
 FEATURES
 source
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 gene
 1..1923
 /gene="PROF"
 CDS
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB55674.2"
 /db_xref="GI:20521155"
 /translation="MRKKKVFSLVLSAAAISTVALSNPSAGGARNFOLDPKGIQTT
 TDAGFSKQGTGAALVSESNVLPKGLQKLETPANNKLHIQFNGPILERTKO
 OLETKAKILDYIPDYAVIYVEQDVKSTSTIEHVESVEPYLPYRIDPOLPTKAS
 ELVAVALDTRQKNKEVQLAGIEQIAQPAISNDVLYITAKPEYKWNVDVARGVKADY
 AQSSIGLYGCGQIVAVADTGLDTRNDSDSHEAFKPTIYALYALGTNNANDTNGHT
 HVAQSVLNGSTNGKMAPQANLVFQSDMSGGGLGPNLQTLFQSAVSAGARIHTN
 SWGAANVAGATTDNRVDYVRKNDMTILFAAGNEGNGGTISAPGTAKNAITVGATE
 NLRPSFGSYADNINHVAQFSRGPDKGRIPKPDVMAPGTFLSARSLADIPGSFWANH
 DSKYAVMGCTSMATPIVAGNVAOLRHFVKNGITTPKSLKALAGIADIGLYPN
 GNCQAGVTLDKSLNAVYVNESLSSTQKATYSFTATAGKPLKISLVWSDAPASTTA
 SVTLNDLDLVITAPNGCTQYVGNDFTSYNDNDGNNVENVFINAPQSGTITIEVQA
 YNPVFGQTFSLAIVN"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,31e-125 Length: 1923
 Score: 2125.50 Matches: 406
 Percent Similarity: 97.93% Conservative: 19
 Best Local Similarity: 93.55% Mismatches: 8
 Query Match: 94.36% Indels: 1

DB: 1 Gaps: 1

US-09-985-689A-7 (1-433) x AB051423 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPhcGlyLeuTyr 20

Db 619 AATGATGTTGGCGTGGAAATGTCACAGCGGATGGCTCAGAGCAGTACCGGTTGTAT 678

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 679 GGACAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTGCATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 739 TCGATGATGAAGCTTCCGCGGAAATTTACTCCATTTATGATGATGGAGCGAGCAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79

Db 799 AATGCCAATGATACCAATGGTGTACGATGCGATGCGTGGCTCCGTTAGGAAACGGC 858

Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99

Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCCAATCTATCATGATAGC 918

Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119

Db 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGAAACCTTATTCAGCCAGCATACAGT 978

Qy 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139

Db 979 GCTGTGTCAGAAATTCATAAACTCTCGGGAGCAGCAGTCAATGGGCTTACACACA 1038

Qy 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159

Db 1039 GATTCCAGAAATGGATGACTATGTGCGCAAAATGATATGACGATCTTTTCGTGCC 1098

Qy 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAlaIle 179

Db 1099 GCGAATGAAGGACCGGAAACCATCAGTCAGCCAGGACAGCTAAATAATGCAATA 1158

Qy 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199

Db 1159 ACAGTCGGAGCTACCGAAACCTCCGCCAAGCTTTGGGTCTTATCGGCAAAATATCAAC 1218

Qy 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219

Db 1219 CATGTGGCACATCTCTTCACGTGGACCGCAAGGATGAGCGATCAACCGCATGTC 1278

Qy 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239

Db 1279 ATGGCACCGGAAACGTTTCATATCAGCAAGATCTTCTTTCGCGGATTCCTCCTTC 1338

Qy 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyIleThrSerMetAlaThrProile 259

Db 1339 TGGCGAACCATGACAGTAATATGATACATGCGGTGGACGTCCATGGCTACACCGATC 1398

Qy 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279

Db 1399 GTTGTGGAACCGTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGAGCATCACACA 1458

Qy 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299

Db 1459 AAGCCTTCCTCTATAAAGCGGACCTGATGTCGCGTGCAGCTGACATCGGCTTGGCTAC 1518

Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319

Db 1519 CCGAACCGTAACCAAGGATGGGACGAGTGACATTCGATAAATCCCTGAACGTTGCCTAT 1578

Qy 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339

Db 1579 GTGAACGAGTCCAGTCTCTATCCACGCGCAAAAGCGAGCTACTTCGTTTACTGCTACT 1638

Qy 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359

Db 1639 GCCGCAAGCCTTTGAAAATCTCCCTGTATGGTCTGTATGCCCTCGGACACACTGCT 1698

Qy 360 SerLeuThrLeuValAlaAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379

Db 1699 TCCGTAACGCTTGTCAATGATCTGACCTTGTCTATTACCGTCCAAATGGCACACAGTAT 1758

Qy 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399

Db 1759 GTAGAAATGACTTTTACTTCGCCATACCAATGATAACTGGGATGGCCGCAATAACGTAGAA 1818

Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419

Db 1819 AATGATATTATTATATGACCAACAAAGCGGACGTATACATATGAGGTACAGGCTTATAAC 1878

Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433

Db 1879 GTACCGGTGGACACAGACCTTCTCGTTGGCAATTGGAAAT 1920

RESULT 6

AR368117 1923 bp DNA linear PAT 12-SEP-2003

LOCUS AR368117

DEFINITION Sequence 5 from patent US 6376227.

ACCESSION AR368117

VERSION AR368117.1 GI:34601778

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1923)

AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y., Shikata,S. and Nomura,M.

TITLE Alkaline protease

JOURNAL Patent: US 6376227-A 5 23-APR-2002;

FEATURES

Location/Qualifiers

1..1923

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores: 1,31e-125 Length: 1923

Pred. No.: 2125.50 Matches: 406

Score: 2125.50

Percent Similarity: 97.93% Conservative: 19

Best Local Similarity: 93.55% Mismatches: 8

Query Match: 94.38% Indels: 1

DB: 6 Gaps: 1

US-09-985-689A-7 (1-433) x AR368117 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPhcGlyLeuTyr 20

Db 619 AATGATGTTGGCGTGGAAATGTCACAGCGGATGGCTCAGAGCAGTACCGGTTGTAT 678

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 679 GGACAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGTGCATGACAGT 738

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 739 TCGATGATGAAGCTTCCGCGGAAATTTACTCCATTTATGATGATGGAGCGAGCAAT 798

Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79

Db 799 AATGCCAATGATACCAATGGTGTACGATGCGATGCGTGGCTCCGTTAGGAAACGGC 858

Qy 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99

Db 859 TCCACTAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCCAATCTATCATGATAGC 918

Qy 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119

Db 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGAAACCTTATTCAGCCAGCATACAGT 978

QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThr 139
Db 979 GCTGTGTCAGAAATTCATCAAACTCCTGGGAGCAGCAGTGAATGGGCTTACACACA 1038
QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
Db 1039 GATTCAGAAATGTGGATGACTATGTGGCCAAAATGATATGACGATCCTTTTCGCTGCC 1098
QY 160 GlyAsnGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGAATGAAGGACCGAACCGCGGAACCATCAGTCAGCCAGGACAGCTAAATAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGTCTTATGGGACAATATCAAC 1218
QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCAGCTTCTCTTCAGTGGACCGCAAGAGGATGGACGATCAAAACCGGATGC 1278
QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCAGGAACTCATATCATACAGAGATCTTCTTTCGACCGGATTCCTCTTC 1338
QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
Db 1339 TGGCGAACCATGACAGTAAATATGATACATACATGCGGTGGACGTCATCGCTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
Db 1399 GTTGTCTGGAACCGTGGCACAGCTTCGTGAGCATTTTGTGAAACACAGAGGATCACACCA 1458
QY 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db 1459 AAGCCCTCTCTATTAAACCGGACCTGATTCGCGGTGAGCTGACATCGGCTTGGCTAC 1518
QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db 1519 CCGAACGGTAACCAAGGATGGGAGCAGTGATCATTTGGATTAATCCTGTAACGTTGCCTAT 1578
QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339
Db 1579 GTGAACGATGCTTCTATCCACCGACCTGATTCGCGGTGAGCTGACATCGGCTTGGCTAC 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359
Db 1639 GCGGCAAGCCTTGAATCTCCCTGGTATGCTGATGCTGCTGCGGAGCACACTGCT 1698
QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379
Db 1699 TCGTAACGCTTGTCAATGATCTGGACCTTGTCAATACCGCTCCCAATGGCACACAGTAT 1758
QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyValArgAsnValGlu 399
Db 1759 GTAGGAAATGACTTTACTTCGCCATACATGATACTGGATGGCGGCAATAAGCTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
Db 1819 AATGTATTATTAAATGCCACCAAGCGGAGCGTATACAAATGAGGTACAGGCTTATAAC 1878
QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 1879 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTTGGAAT 1920

RESULT 7

AR368118 LOCUS 1923 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 7 from patent US 6376227.
ACCESSION AR368118
VERSION AR368118.1 GI:34601779
KEYWORDS Unknown.

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 1923)

AUTHORS Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kageyama,Y., Shikata,S. and Nomura,M.

TITLE Alkaline protease

JOURNAL Patent: US 6376227-A 7 23-APR-2002;

FEATURES Location/Qualifiers

source

1..1923

/organism="unknown"

/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.:	2,71e-125	Length:	1923
Score:	2120.50	Matches:	405
Percent Similarity:	97.70%	Conservative:	19
Best Local Similarity:	93.32%	Mismatches:	9
Query Match:	94.16%	Indels:	1
DB:	6	Gaps:	1

US-09-985-689A-7 (1-433) x AR368118 (1-1923)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 619 AATGATGTTGGCGTGGAAATGTCAACCGGATGTGGCTCAGACGACTACGGGTTGTAT 678
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GCACAAGACACATCGTAGCGTTGCCGATACAGGGCTTGATACAGGTGCGAATGACAGT 738
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATGCGATGAAGCCCTCCGCGGAAATTTACTGCATTTATGATTTGGGACGACCAAT 798
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn --- 79
Db 799 AATGCCAATGATACGAATGGTCATGTCAGCATGTGGCTGGCTCCGTATTAGGAAACGGC 858
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 859 TCCACTAATAAAGGAATGGCGCTCAGCGAATCTAGTCTTCCAACTATTCATGCGATGC 918
QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
Db 919 GGTGGGGACTTGGAGGACTACCTTCGAATCTGCAAACTTATTTCAGCAAGCATAAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThr 139
Db 979 GCTGTGGCAGAAATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACACA 1038
QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
Db 1039 GATTCAGAAATGTGGATGACTATGTGGCAAAATGATATGATGACATCTTTTCGCTGCC 1098
QY 160 GlyAsnGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db 1099 GGAATGAAGGACCGAACCGCGGACCATCATGTCACCGACACAGCTTAATAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
Db 1159 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGTCTTATGGCGACAATATCAAC 1218
QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
Db 1219 CATGTGGCAGCTTCTCTTCAGTGGACCGCAAGGATGGACGATCAAAACCGGATGC 1278
QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db 1279 ATGGCACCAGGAAACGTTCTATCATACAGAGATCTTCTTTCGACCGGATTCCTCTTC 1338
QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259

QY 240 TtpAlaAsnHisAspSerIysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 259
 DB 721 TGGGCAAAACCATGACAGCAAAATATCTTATATGGGTGGAAATCATCCATGGGACACCGGAT 780
 QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAspGlyValThrPro 279
 DB 781 GTAGCTGGTAAAGTTCACAGTACGTGAACATTCATCAAAAACAGAGMAATCCTCT 840
 QY 280 LysProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
 DB 841 AAACCATCTCTGCTGAAAGCAGCTCTTATTCGCGAGCAACTGATATCGGTCTTGGCTAT 900
 QY 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
 DB 901 CCGAGTGGAAACCAAGGATGGGAAGAGTAACTTGGACAACTTATTAATGAGCTTTC 960
 QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339
 DB 961 GTAAATGAACAAGCTCTTATCTACTAACCACAAAGGCTAGCTATTCATTACTGCACAA 1020
 QY 340 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProGlySerThrThrAla 359
 DB 1021 TCAGGCACAACTTTCAGAGATTCATTTGGTTCGATCGCCGCAAGTACTTCGCCA 1080
 QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValLleThrAlaProAsnGlyThrLysTyr 379
 DB 1081 TCGGTATCATTTGGTGAATGATCTGGATCTGGTGTATACAGCTCCAAATGGAACAAAGTAT 1140
 QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnValGlu 399
 DB 1141 GTTGGAAACGATTTACTTCTCTCCCTATGATATTAATTAACCTGGATGACCTACATGTAGAG 1200
 QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsn 419
 DB 1201 AACGTGTTTATCAATGCTCCGCAAGCGGAACGTATACAGTTCAGGTAGGTACAGGCTTACAAT 1260
 QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
 DB 1261 GTTCACAAAGGCCCGCGCGCTTTCTTTGGCTATTGTGAAC 1302

RESULT 9
 AB046403
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS

TITLE
 1920 bp DNA linear BCT 10-MAY-2002
 Bacillus sp. 9860 PROA gene for protease, complete cds.

AB046403.2 GI:20521152
 Bacillus sp. 9860
 Bacillus sp. 9860
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
 Horikoshi, K.
 Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail:387.185@kamatanet.kao.co.jp,
 Tel:81-285-68-7400, Fax:81-285-68-7403)
 On May 9, 2002 this sequence version replaced gi:12381938.
 Location/Qualifiers
 1. 1920
 /organism="Bacillus sp. 9860"
 /mol_type="genomic DNA"
 /strain="9860"

Gene
 CDS
 /db_xref="taxon:133778"
 1. 1920
 /gene="PROA"
 1. 1920
 /gene="PROA"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB21266.2"
 /db_xref="GI:20521153"
 /translation="MRKKVFLSVLSAAAILSTVALNPSAGDARTFDLDFKGIQTWT
 DVSGFSKQRTGAAFLVSEGNKVLKGLLKLLETVPANNKLHIVQNPDLIFETKQK
 LETTGAKILDYIPDYAVIVEGDVQSVKESIHEVSVEPYLPKYKIDPLFKGAST
 LVKALADTKNNKEVOLRGIESIAOVASNDVHYLTAKPEYKVMNDVARGIVKAQVA
 OSSVGLVGGQGVAVADTGLDTGRNDSMEAEFRGKITALYALGRNNDNTNGHGT
 VAGSVNGATYTRSDVDYVRKNDMTILFAAGNPNNGTISAPGAKNAITVGATENS
 WGSVNGAYTTDSRNDYVRKNDMTILFAAGNPNNGTISAPGAKNAITVGATENS
 LRPFSGYANINHVAGNSRGPTKGRIPKDYNAFTVLSARSLADSPGFWANHD
 SKYMGCTGMATIVAGNVQALREHFVKNGITPKPSLLKALIAAGADVGLGYGNG
 NQMGRTVLDKSLNVAYNSSSALSTQKATYTTATAGKPLKISLWSDAPASTAS
 VTLVNDLDELITAPNGTRYVNDFSAPFDNNWDGRNVENVFINSFQSGTYTTEVQAY
 NVVPVQPNFSLIYN"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.16e-124 Length: 1920
 Score: 2110.50 Matches: 404
 Percent Similarity: 97.24% Conservative: 18
 Best Local Similarity: 93.09% Mismatches: 11
 Query Match: 93.72% Indels: 1
 DB: 1 Gaps: 1

US-09-985-689A-7 (1-433) x AB046403 (1-1920)

QY 1 AsnAspValAlaAspGlyIleValLysAlaAspValAlaGlnAsnAsnAspGlyLeuTyr 20
 DB 616 AATGATGCGCCAGAGTATTGTCAAAGCGGATGTGGCAGAGCAGTACGCTTTGTAT 675
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 676 GGACAGGCCAGATTGTCGCGAGTTGCGGATACGATGATGATGATGATGATGATGATGAT 735
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 DB 736 TCGATGATGAAGCCTTCCGCGGTAAATAACAGCATTATATGCTACGCTGGTGGACGAAT 795
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
 DB 796 AATGCGAATGATACGAACCGGTATGTTGACCCATGTGGCAGGTTCCGCTATTAGGAATGCG 855
 QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
 DB 856 GCAACGAATAAGCAATGGCACCTCAAGCGAATCTGGTTTTCATCCATCATGATGATGC 915
 QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
 DB 916 AGTGGTGGCTTGAGGCTTGCCCTTCCAATCTCAACACCTTATTCAGCCAGCATTCAGT 975
 QY 120 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValProValAsnGlyAlaTyrThr 139
 DB 976 GCAGGTGCCAGATTATCAAACTCTCTGGGGGCGAGCGGTGAATGGGGCTACACGACA 1035
 QY 140 AspSerArgenValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
 DB 1036 GATTCCAGAAATGGAATGACTATGTAAAGGAAAATGATATGACGATTCTTTTCGGCGCT 1095
 QY 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
 DB 1096 GCGAATGAAGCGCAACGCGGCTACCATCAGTGCACCTGGTACGCTAAAACCGCCATA 1155
 QY 180 ThrValGlyAlaThrGluAsnLeuAcgProSerPheGlySerTyrAlaAspAsnIleAsn 199
 DB 1156 ACAGTGGCGGCACACCGAAAACCTGCGTCCAAAGCTTCGGTTCCTATGACGATAATATTAC 1215

QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
DB 1216 CACGTTGCACAGTTCTCTTCCCGTGGCCGACAAAGATGGCGAATCAAGCCCTGATGTC 1275
QY 220 MetAlaProGlyThrTyrlleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
DB 1276 ATGGCGCCAGGACATACATTTTATCAGCAGATCTTCTTGCACCGGATTCTCTCTTC 1335
QY 240 TrpAlaAsnHisAspSerLysTyrlaTyMetGlyGlyThrSerMetAlaThrProIle 259
DB 1336 TGGCGGAATCATGACAGCAAAATATGCCCTATATGGGTGGAACTGCCTCAGCAACCGGATT 1395
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
DB 1396 GTTGGCGGGAATGTTGCACAGCTCGTGAGCATTTTGTGAAAATAGAGGAATCACTCT 1455
QY 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
DB 1456 AAGCCTTCCCTATTGAAAGCAGCTTTGATTGCGGTGCTGTGATTTGGATTGGGTAT 1515
QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
DB 1516 CCGAACGGAACCAAGATGGGCCAGTACCTTGGTAAATGTTGGAACGTTGCCAT 1575
QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrlleSerPheThrAlaGln 339
DB 1576 GTGAACGGAATCCAGTGCCTATCAACTAGCCAAAAGCGCATATACCTTACTGCAACG 1635
QY 340 AlaGlyLysProLeuLysAlaSerLeuValTrpSerAspAlaProGlySerThrAla 359
DB 1636 GGGGCAAGCCATTGAAATCTCCCTGTGTGTCGGATGCCCTGCAAGCACTACTGCT 1695
QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrl 379
DB 1696 TCTGTAACCTGTCAATGATTGGATTGGTCAATTAGCACCACCAACGGAACAGATAT 1755
QY 380 ValGlyAsnAspPheThrAlaProTyrlleAspAsnAsnTrpAspGlyArgAsnValGlu 399
DB 1756 GTCGGGAATGATCTTCCAGCACCATTTGCATAAATCGGATGGCGGCAATAAGCTAGAA 1815
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrlleValGluValGlnAlaTyrlAsn 419
DB 1816 AATGATTATTAAATTCGCCCAAGTGAACATATACCATGAGTGCAGCATATAAT 1875
QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
DB 1876 GTGCCGGTTGGACCAAACTCTCGTTGGCAATTGTGAAC 1917
RESULT 10
AR368116
LOCUS AR368116 1920 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 3 from patent US 6376227.
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1920)
TAKAIWA,M., OKUDA,M., SAEKI,K., KUBOTA,H., HITOMI,J., KAGEYAMA,Y.,
SHIKATA,S. and NOMURA,M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores: 1.16e-124 Length: 1920
Pred. No.: 2110.50 Matches: 404
Score:

Percent Similarity: 97.24% Conservative: 18
Best Local Similarity: 93.09% Mismatches: 11
Query Match: 93.72% Indels: 1
DB: 6 Gaps: 1
US-09-985-689A-7 (1-433) x AR368116 (1-1920)
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyrl 20
DB 616 AATGATGTGCCAGAGGATTTGTCAAAGCGGATGTGGCAGAGCAGCTACGGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 676 GCACAGGCGCAGATTGTCCAGTGTCCGATCTGCGATCTGCGATGTGATGAGGAGAAACACAGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrlAlaLeuGlyArgThrAsn 60
DB 736 TCGATGTCATGAGCCTTCCCGGTAAATAACAGCACTATATGCACTGGTGGGACGAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn-- 79
DB 796 AATGCGAATGATACGAACGTCATGTTACCCATGTGGCAGGTTCGGTATTAGGAATGCG 855
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
DB 856 GCACCGAATAAAGGAATGCACTCAAGCGAATCTGGTTTTCAATCCATCATGGATAGC 915
QY 100 GlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrlSer 119
DB 916 AGTGGTGGGCTTGGAGGCTTGCCTTCCAATCTGCAAACTTATTAGCAAGCATTCACT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrlThr 139
DB 976 GCAGGTGCCAGAATTCATCAAACTCCTGGGGGGCAGCGTGAATGGGGCCTACAGCA 1035
QY 140 AspSerArgAsnValAspAspTyrlValArgLysAsnAspMetThrIleLeuPheAlaIle 159
DB 1036 GATTCCAGAAATGTGGATGACTATGTAGAGAAAATGATATGACGATTCTTTTCGGGCT 1095
QY 160 GlyAsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
DB 1096 GGGAAATGAAAGCGCAACCGCGGTACCATCAGTGACCTGGTACGGCTAAAACCGCAT 1155
QY 180 ThrValGlyValaThrGluAsnLeuArgProSerPheGlySerTyrlAlaAspAsnIleAsn 199
DB 1156 ACAGTGGCGCAACCGAAACCTGCTCCAGCTTCGGTTCCTATGACAGTAATAATTAAC 1215
QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
DB 1216 CACGTTGCACAGTTCTTCCCGTGGCCGACAAAAGATGGGCGAATCAAGCCTGATGTC 1275
QY 220 MetAlaProGlyThrTyrlleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
DB 1276 ATGGCGCCAGGACATACATTTTATCAGCAGATCTTCTTTCGACCCGATTCCTCTTC 1335
QY 240 TrpAlaAsnHisAspSerLysTyrlAlaTyMetGlyGlyThrSerMetAlaThrProIle 259
DB 1336 TGGCGGAATCATGACAGCAAAATATGCTATATGGGTGGAACTGCCTCAGCAACCGGATT 1395
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
DB 1396 GTTGGCGGGAATGTTGCACAGCTCGTGAGCATTTTGTGAAAATAGAGGAATCACTCT 1455
QY 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
DB 1456 AAGCCTTCCCTATTGAAAGCAGCTTTGATTGCGAGTGTCTGATGTTGGATTGGGTAT 1515
QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
DB 1516 CCGAACGGAACCAAGATGGGGCCAGTACCTCGATGAATAATCGTTGAGCTGCTAT 1575
QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrlSerPheThrAlaGln 339

```
Db 1576 GTGAACGAATCCAGTGCCTTATCAACTAGCCAAAAGCGACATATACCTTTACTGCAACG 1635
Qy 340 AlGlyYsProLeuYsIleSerLeuValTrpSerAspAlaProGlySerThrThrAla 359
Db 1636 GCGGGCAAGCATTGAAATCTCCCTGGTATGTCGATGCGCCCTCCAGACACTACTGCT 1695
Qy 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrIleYr 379
Db 1696 TCTGTAACTTCTGCAACCACTTTCGATTTGGTTCATTACAGCACCAAAACGAAACATAT 1755
Qy 380 ValGlyAsnAspPheThrAlaProTyZAspAsnAsnTrpAspGlyArgAsnValGlu 399
Db 1756 GTGCGGAATGACTTCTCAGCAACCACTTTCGATTTGGTTCATTACAGCACCAAAACGAAACATAT 1815
Qy 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyZThrValGluValGlnAlaTyZAsn 419
Db 1816 AATGTATTATTATTTCGCCCAAGTGGCAATATATACCATTTAGGTGGCAAGCATATAT 1875
Qy 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 1876 GTGCGGGTGGACCAAACTTCTCGTTGGCAATTGTGAAC 1917

RESULT 11
AB046405 1299 bp DNA linear BCT 23-JAN-2001
LOCUS Bacillus sp. SD521 PROD gene for protease, partial cds.
DEFINITION Bacillus sp. SD521
ACCESSION AB046405
VERSION AB046405.1 GI:12381942
KEYWORDS
SOURCE Bacillus sp. SD521
ORGANISM Bacillus sp. SD521
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (sites)
AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horikoshi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE 20568675
PUBMED 11118284
REFERENCE 2 (bases 1 to 1299)
AUTHORS Saeki,K.
DIRECT SUBMISSION
TITLE Submitted (20-JUL-2000) Katsuhisa Saeki, Kao Corporation,
JOURNAL Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES
source
1..1299
location/Qualifiers
organism="Bacillus sp. SD521"
mol_type="genomic DNA"
strain="SD521"
/db_xref="taxon:133780"
1..1299
/gene="PROD"
<1..>1299
/gene="PROD"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21268.1"
/db_xref="GI:12381943"
/translation="NDVARGIVKADVQAQNNYGLYGGQVAVADVADTGLDTRNDSSMHE
AFRGKITALYALGRNTNANDPNGHGVHAGSVLGNALNKGMAPQANLVFQSIIMPSSGG
LGLPSNTLSQANNGARLHTNSWGA PVNGAYTANSROVDEVVRNNDMTLVLPFAAG
NEGPNSTISAGTKAATVATENRPSFGSLADNPNHIAQFSRSGATRDGRPKPD
VTAPGFFILASRSSAPDSSFWANTNSKATYMGTSMTAIPVAGNVQAQIRHFINKRG
ITPKESLIKAALLIAGTADVLGPGDQGWRTLDKSLNVAYNVEATLATGQATY
SFQAQAGKPLKSLVMTDAPGTSTATYLVNDLDLVITAPNGQKTVGNDFSYPYDNNW
DGRNNVENVFINAQSGTITIEVQAVNNVPSQPRFSLAIH"
```

```
Alignment Scores: 1.16e-120 Length: 1299
Pred. No.: 2044.00 Matches: 386
Score: 2044.00 Conservative: 24
Percent Similarity: 94.69% Mismatches: 23
Best Local Similarity: 89.15% Indels: 0
Query Match: 90.76% Gaps: 0
DB: 1
US-09-985-689A-7 (1-433) x AB046405 (1-1299)
Qy 1 AsnAspValAlaAArgGlyIleValIysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 1 AATGTGTAGCAAGAGGAAATAGTAAACACAGACGTTGCACAAAACAATATCGGACTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaAAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GCGAAGGTCAAGTAGTTGCGAGTAGCGGATACGGGTTTAGATACAGGTCGTAACGATAGT 120
Qy 41 SerMetHisGluAlaPheAArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGTCATGAGCAATTCGTGGGAAATATACAGCTCTTTACGCGTTAGGAAGAACTAAC 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 181 AATGCAATATGATCCGAATGGGATGTTAGCATGTAGCAGGTTCTGTACTTGGTAATGCT 240
Qy 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
Db 241 TTAATAAAGGAATGGCTCCGACAGCTACTTAGTCTTCCATCTATTATGATAGCAGC 300
Qy 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
Db 301 GCGAGGATTAGTGGATTCACATCGAATTTGAATAGCTTATTATTAGTCAAGCTTGGAAATGCT 360
Qy 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrThrAsp 140
Db 361 GCGGCTAGAAATTCATACTAACTCTTGGGTGCTCCAGTAAATGGAGCGTACTCTGCTAAC 420
Qy 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160
Db 421 TCGAGACAAGTGGATGAGTATGTTGGAATAATATGATATGACGGTACTTTTTTCAGCAGGT 480
Qy 161 AsnGlyGlyProGlySerGlyThrIleSerAlaProGlyThrAlaIysAsnAlaIleThr 180
Db 481 AATGAAGAGTCTTAATTCAGGAACAATTAGTGTCTCAGGCACAGCCGAAAAATGCCATTAGC 540
Qy 181 ValGlyValaThrGluAsnLeuArgProSerPheGlySerTyZAlaAspAsnIleAsnHis 200
Db 541 GTCGGCGCAACGGAAAACTATCGCCCGAGCTTCGGTTTCATTAGCAGATAACCCCAATCAT 600
Qy 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleIysProAspValMet 220
Db 601 ATCGCACAATTTTCATCAAGAGAGGAGCTACGAGAGATGGACGAATTAACACGACGTAACA 660
Qy 221 AlaProGlyThrTyZThrIleLeuSerAlaAArgSerSerLeuAlaProAspSerPheTrp 240
Db 661 GCTCCTGGAAACATTTATTTTATCAGCAGCTTCTCTTAGCCCCCAGACTCTTCGTTTGG 720
Qy 241 AlaAsnHisAspSerTyZThrAlaTyZMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 721 GCGAATTATAACAGTAGTAGTATGCGTACATCGGCGGTACTCTTATGCGCAGACACCTAGATT 780
Qy 261 AlaGlyAsnValAlaGlnIleuArgGluHisPheValIysAsnArgGlyValThrProLys 280
Db 781 GCGGGGAATGTGCGGCAATTTACGCGAGCAATTTATATAAATAAGAGGAATATACACCTAAA 840
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 841 CCTTCTTAATAAAGCTGCACCTTATCGCTGGGCTACTCATGCTTGGTTGGTTAGGATATCCA 900
Qy 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
Db 901 AGTGGTGCACCAAGGCTCGGGCGGTGTACTCTAGATAAATCGTTAAATAGCGTAIGTC 960
```

```
Qy 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaAlaThrTyrSerPheThrAlaGlnAla 340
Db 961 AATGAAGCAATCGCATTTAGCAACAGCAAAAAGCAACGTTATTCGTTCACAGCAACAGCG 1020
Qy 341 GlyLysProLeuLysLysSerLeuValTyrSerAspAlaProGlySerThrAlaSer 360
Db 1021 GGTAAACCTTTAAANAATCTGTTAGTATGACAGATGACCTGGAAGTACAACTGCATCT 1080
Qy 361 LeuThrLeuValAsnAspLeuAlaProGlyLysLeuValThrAlaProGlyThrLysTyrVal 380
Db 1081 TATACACTAGTTAATGATTTAGATCTAGTTATTTACTCTCCGATGACAAAAATATGTA 1140
Qy 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTyrAspGlyArgAsnValGluAsn 400
Db 1141 GGAAATGATTTTATGTTATCTTATGATATTAACCTGGATCGTCCGCAACAATGTTGAGAC 1200
Qy 401 ValPheLeuAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnVal 420
Db 1201 GTATTTATAACGCTCCGCAATCTGGAACGTATACAAATTCAGGTTCAAGCGTTAATGTA 1260
Qy 421 ProValSerProGlnThrPheSerLeuAlaLeuValHis 433
Db 1261 CCATCTGGCCACACGCTTCTCACTAGCTATCGTACAT 1299

RESULT 12
AB046402
LOCUS AB046402 1299 bp DNA linear BCT 16-JUN-2001
DEFINITION Bacillus sp. D6 PROA gene for protease, partial cds.
ACCESSION AB046402
VERSION AB046402.1 GI:12381936
KEYWORDS
SOURCE
ORGANISM Bacillus sp. D6
REFERENCE Bacillus sp. D6
AUTHORS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
Saeki,K.; Okuda,M., Harada,Y., Kobayashi,T., Ito,S., Takami,H. and
Horikoshi,K.
TITLE Novel oxidatively stable subtilisin-like serine proteases from
alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
evolutionary relationships
Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
20568675
11118284
2 (bases 1 to 1299)
Direct Submission
Saeki,K.
Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:387185@kasanet.kao.co.jp,
Tel:81-285-68-7400, Fax:81-285-68-7403)
Location/Qualifiers
1..1299
/organism="Bacillus sp. D6"
/mol_type="genomic DNA"
/strain="D6"
/db_xref="taxon:127889"
1..1299
/gene="PROA"
<1..>1299
/gene="PROA"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAB21265.1"
/db_xref="GI:12381937"
/translation="NDVARGIVKADYQNNYGLYKQGVAVADTGLDTGRNDSMHE
AFRGKITALYALGRNANDPNHGTHVAGSVLGNALNKGMAPQANLVFQSDSSGG
LGLPSNLNLFQAWNAGARLHTNSWGPVNGVAYTANRSQVDEYVRNNDMTLVFAAG
NEGPNSTISAPCTAKNALTVGATENYRPSFGSIADNPNIHQFSSRGATRDGRKPD
VTAPGTFILSSSLAPDSSFWANYSKYVMGTSMATPIVAGNVAOLRHFHKNRG
ITPKSLIKALLAGATDVLGVPSGDQGWGRVTLDKSLNVAVNEATLTGOKATY
SFQTQAGKFLKLSLWTDAPGTSATYTLVNDLVLITAPNGQKIVGNDFSYPYDNNW
```

FEATURES

```
source
gene
CDS
```

ORIGIN

```
Alignment Scores:
Pred. No.: 2,08e-120 Length: 1299
Score: 2040.00 Matches: 385
Percent Similarity: 94.46% Conservatives: 24
Best Local Similarity: 88.91% Mismatches: 24
Query Match: 90.59% Indels: 0
DB: 1 Gaps: 0

US-09-985-689A-7 (1-433) x AB046402 (1-1299)

Qy 1 AsnAspValAlaAlaArgGlyLysValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 1 AATGATGTAGCAAGAGGAATAGTAAAGACAGACGTTGCACAAAACAAATTCGGACTATAT 60
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 61 GGCAAGGTCAGTAGTTGTCAGTAGCGATAGCGGTTTATAGTACAGGTCGTCAGATAGT 120
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 121 TCTATGCGATGAGCATTCGGTGGGAAATTAACAGCTCTTTACGGCTTAGGAAGAATAAC 180
Qy 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 181 AATGCAAAATGATCCGAATGGGATGCTAGCTAGCTGGTTCGTGCTTGGTATGCT 240
Qy 81 ThrAsnLysGlyWetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
Db 241 TTAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAATCTATTATGATAGCAGC 300
Qy 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
Db 301 GGAGGATTAGGAGGATTCACATCGAATTTAAATACGTTATTATTAGTCAAGCTTGAATGCT 360
Qy 121 GlyAlaAlaTyrIleHisThrAsnSerTyrGlyAlaProValAsnGlyValAlaTyrThrAsp 140
Db 361 GGCGCTAGAAATTCATACCTTAACCTTTGGGGGGCCCCAGTAAATGGAGCGTACACTGAAC 420
Qy 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGly 160
Db 421 TCGAGCAAGTGGATGAGTATGTTGCAACAATGATATGACGGTACTTTTTCAGCTGGA 480
Qy 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 481 AATGAAGGACCTTAACCTTCGAAACAATAGCGCTCCAGGACAGCGAATAATCCATTAGC 540
Qy 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
Db 541 GTCGGCGCAACGGAAACCTACCGCAAGTTTGGTTCAATTCAGATTAACCTTAATCAT 600
Qy 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyValArgLysProAspValMet 220
Db 601 ATCGCACAATTTTTCATCGAGAGGAGGTACGAGAGATGGAGCAATTAACACAGACGTAA 660
Qy 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
Db 661 GCTCCTCGAACATTTATATTATTCAGCAGCGCTCTTCTTTAGCACAGACTCTTCGTTTGG 720
Qy 241 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 721 CGCAATTTATACAGTAAATGCTATGCTATGCGGCGGTACCTCTATGCGGACACCTATAGT 780
Qy 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
Db 781 CGCGGGGATGTCGGCAATTTACCGGAGCAATTTTATAAAAAATAGAGAAATTACACCTAAA 840
Qy 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 841 CCTTCTCTTAATAAAGCTGCACTTATCGTGGGGGCTACTGATGTTGTTAGGATATATCCA 900
```

QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
 Db 901 AGTGGTGACCAAGCGTGGGGCGGTACTTTAGATTAATCGTTAAATGTAGCGTATGTC 960
 QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTySerPheThrAlaGlnAla 340
 Db 961 AATGAAGCAACTGTCATTAACAACAGACAAAAGCAACGTATTCCTCCAAACGCAAGCG 1020
 QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrThrAlaSer 360
 Db 1021 GGTAAACCACTTAAATCTCGTTAGTATGGACAGATGCACCTGGAGTACACAGCATCT 1080
 QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyVal 380
 Db 1081 TATACACTAGTTAATGATTTAGTCTAGTATTTACTGCTCCGAATGGCAAAAATATGTA 1140
 QY 381 GlyAsnAspPheThrAlaProTyAspAsnAsnTrpAspGlyValGlnAsnValGlnAsn 400
 Db 1141 GGTAAATGATTTAGTTATCTTATGATTAATTTGGATGGTTCGCAACAATGTTGAGAAC 1200
 QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyAsnVal 420
 Db 1201 GTATTTATAACGCTCCGAATCTGGAACGTATACAATTGAGGTTCAAGCGTATAACGTT 1260
 QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
 Db 1261 CCATCTGGACCAACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 13

AB046404
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (sites)
 Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
 Horikoshi, K.
 Novel oxidatively stable subtilisin-like serine proteases from
 alkaliphilic bacillus spp.: enzymatic properties, sequences, and
 evolutionary relationships
 Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 20568675
 11118284
 2 (bases 1 to 1299)
 Saeki, K.
 Direct Submission
 Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
 Tochigi 321-3497, Japan (E-mail: 387185@kasanet.kao.co.jp,
 Tel: 81-285-68-7400, Fax: 81-285-68-7403)
 Location/Qualifiers
 1. .1299
 /organism="Bacillus sp. Y"
 /mol_type="genomic DNA"
 /strain="Y"
 /db_xref="taxon:133779"
 1. .1299
 /gene="PROC"
 <1. .>1299
 /gene="PROC"
 /codon_start=1
 /evidence="experimental
 /transl_table=11
 /product="protease"
 /protein_id="BAB21267.1"
 /db_xref="GI:12381941"
 /translation="NDVARGIVKADVAGNQNYGLYGGQGVAVADTGLDGRNDSWHE
 AFRKITALYALGRTNASDPNGHGTTHVAGSLGNALNKGAPQANLVFQSIIMDSGG
 LGGLPSNLNTLPSQAWNAGARLHTNSWGPVNGAVTANSRQVDEYVRNDWTVLFRAG

FEATURES

source
 gene
 CDS

ORIGIN

Alignment Scores:
 Pred. No.: 5,77e-120 Length: 1299
 Score: 2033.00 Matches: 384
 Percent Similarity: 94.46% Conservative: 25
 Best Local Similarity: 88.68% Mismatches: 24
 Query Match: 90.28% Indels: 0
 DB: 1 Gaps: 0
 US-09-985-689A-7 (1-433) x AB046404 (1-1299)

QY 1 AsnAspValAlaAspGlyLysValIleValIleValAlaGlnAsnAsnPheGlyLeuTy 20
 Db 1 AATGATGTAGCAAGAGGATAGTAAAGCTGATGTTGCACAAACAACTATTCGGATTATAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 CGACAAGGTCAAGTAGTTGCAGTAGCGACACAGGCTTAGATACAGGTCGTAAAGTAGT 120
 QY 41 SerMetHisGluAlaPheAspGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGTCATGAAGCAATTCGGGGAAATCAAGCTCTTACGGCTTAGGAAGAACTAAT 180
 QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
 Db 181 AATCGAGTGTATCCGAATGGCGATGGCACACATGTAGCAGGTTCTGTACTTGGTAATGCT 240
 QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
 Db 241 TTAATAAAGGAATGGCTCCGCAAGCTAACTTAGCTTCCAACTCTATTATGGATAGCAGC 300
 QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 120
 Db 301 CGAGATTAGGTGGCTTACCACTCGAATTAATAACGTTATTAGTCAAGCTTGAATGCT 360
 QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAsp 140
 Db 361 GGAGCAAGAAATTCATCTACTTCTTGGGGAGCCCAAGTAATGAGGCTACACCTGCTAAC 420
 QY 141 SerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
 Db 421 TCGACACAAGTGGATGAATATGTTTCAAAATATATATATGACCGTACTTTTTCAGCTGCT 480
 QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
 Db 481 AATGAAGTCTCTAATTCAGGAACAATTAGTCTCCAGTACAGCAAAAATGCTATTATCG 540
 QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHis 200
 Db 541 GTGGCGCAACGAAAACATATCGGCCAAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 600
 QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
 Db 601 ATTGCACAATTTTCAATCGAGAGGAGCTACAGAGGATGACAGCAATTAAGCCCTGACGTAACA 660
 QY 221 AlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTy 240
 Db 661 GCTCCTCGGAACATTTATTTTATCAGCAGCTTCTTCTCTTAGCTCCAGACTCTCTCGTTGG 720
 QY 241 AlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 260
 Db 721 GCGAATTTATACGTAATAATCGGTATATGGCGGTACCTCCATGGCGACACCTATTGTT 780
 QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
 Db 781 GCAGGGAATGTCGCGCAATTCAGTGAGCATTTTATAAAAAATAGAGGTATTACTCTCTAAG 840

NEGNSGTISAPGTAKNATVYATENYRPPSGSIADNPNTIAQFSSRGATRDGRKIPD
 VTACTFTLSRSLAPDSSFRANYNSKYAMGSGTSMATPVAAGNVAOLREHFKRG
 TTPKSLIKALILAGTNDVGLGYPNGCGWGRVTLNKLNSLVAYNYEATALATCAKATY
 SFQAGKPLKISLVITWDAPGSITLVNDLDELIVITAPNGQKIVGNDFSYPDNW
 DGRNVENVFINAPQSGTYIIIEVQAYNVPSQFQFSLAIVH"

QY 281 ProSerLeuLeuLysAlaAlaLeuAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 841 CTTCTTTTAAATAAGCTGCACATTATCGTGGTGTACTCATGCTGTTTAGGATATCCT 900
QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
Db 901 AATGGTACCAAGCTGGGGCGGTGTACTCTAAATAAATCGTTAAATGTAGCGGTATGTC 960
QY 321 AsnGluThrSerProLeuSerSerGlnLysAlaAlaThrThrSerPheThrAlaGlnAla 340
Db 961 AATGAACAACTGCATTAGCCACAGACAAAAGCAACGATTTCGTTCCAGCACAAAGCG 1020
QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAlaSer 360
Db 1021 GGTAAACCTTTAAATAATCTCGTTAGTATGACACATGCTCCTGGAAGTAACTGCATCT 1060
QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTrpVal 380
Db 1081 TATACACTAGTTAATGATTAGATCTAGTATTACTGCTCCGATGACAAATATGTA 1140
QY 381 GlyAsnAspPheThrAlaProTyAspAsnAsnTrpAspGlyArgAsnAsnValGluAsn 400
Db 1141 GGAAATGATTTTGTATCTCTTATGATAAATAACTGGATGGTGTGCAACAATGTTGGAAC 1200
QY 401 ValPheIleAsnAlaProGlnSerGlyThrThrValGluValGlnAlaTrpAsnVal 420
Db 1201 GTATTATTAACGCTCCCAATCTGGAACGTATATTAATTGAGTTTCAAGGTATATGTA 1260
QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 1261 CCATCGCCACACGCGTTTCTCACTAGTATCGTACAT 1299

RESULT 14
E03808
LOCUS 2218 bp DNA linear PAT 29-SEP-1997
DEFINITION DNA encoding alkaline protease Ya.
ACCESSION E03808
VERSION E03808.1 GI:2172022
KEYWORDS JP 1992197182-A/1.
SOURCE Bacillus sp.
ORGANISM Bacillus sp.
REFERENCE 1 (bases 1 to 2218)
Tobe,S., Odera,M. and Asai,Y.
TOBE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
PROTEASE YA USING THE DNA
Patent: JP 1992197182-A 1 16-JUL-1992;
JOURNAL LION CORP
COMMENT OS Bacillus sp.
PN JP 1992197182-A/1
PD 16-JUL-1992
PF 28-NOV-1990 JP 1990327110
PI TOBE SEIICHI, ODERA MOTOYASU, ASAI YOSHIO
PC C12N15/57,C11D3/386,C12N9/54,C12N15/57,C12R1:07),(C12N9/54,
PC C12R1:07);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Y strain;
CC *source: clone=PB9A;
FH Key Location/Qualifiers
FH CDS .218..2125
FT /products='precursor of alkaline protease Ya'
FT sig_peptide 218..823
FT FT 824..2122
FT mat_peptide /products='Precursor of alkaline protease Ya'.
FT Location/Qualifiers
1..2218
FT /organism='Bacillus sp.'
FT /mol_type='genomic DNA'
FT /db_xref='taxon:1409'

ORIGIN
Alignment Scores: 1.29e-119 Length: 2218
Pred. No.: 2032.00 Matches: 384
Score: 2032.00
Percent Similarity: 94.46% Conservative: 25
Best Local Similarity: 88.68% Mismatches: 24
Query Match: 90.23% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-7 (1-433) x E03808 (1-2218)
QY 1 AsnAspValAlaAspGlyIleValLysAlaAlaAspValAlaGlnAsnAsnPheGlyLeuTrp 20
Db 824 AATGATGATGACGAGGATAGTAAAGCTGATGTTCACAAAACAATTACGGATTATAT 883
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 884 GGCAAGGTCACCTAGTTGCGACATGCGGACACAGGCTTAGATACAGGTCGTAACGATAGT 943
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTrpAlaLeuGlyArgThrAsn 60
Db 944 TCTATGTCATGAAGCAATTCGCGGGAAATACAGCTCTTTACGCGTTAGGAGAACTAAT 1003
QY 61 AsnAlaAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1004 AATGCGAGTGATCCGAATGGGCATGGCACACATGATGACAGGTTCTGTACTTGGTAATGCT 1063
QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
Db 1064 TTAAATGAAGGAATGGCTCCGCAAGCTAATCTAGTCTTCCCAATCTATTATGGATAGCAGC 1123
QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTrpSerAla 120
Db 1124 GGAGGATTAGTGGCTTACCATCGAATTAATACGTTATTAGTCAAGCTTGAATGCT 1183
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTrpThrAsp 140
Db 1184 GGAGCAAGAATTCACTAATCTTTGGGAGGCCAGTAATATGGAGCGTACACTCCTAAC 1243
QY 141 SerArgAsnValAspAspTrpValArgLysAsnAspMetThrIleLeuPheAlaIleGly 160
Db 1244 TCGAGCAAGTGGATGATGATGTTTCGAATAATGATATGACGGTACTTTTGCAGCTGGT 1303
QY 161 AsnGlyLysProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1304 AATGAAGTCTCAATTCAGGAACAATATGCTCCAGGTACAGCGAAAATGCTATTACG 1363
QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTrpAlaAspAsnIleAsnHis 200
Db 1364 GTCCGCGCAACGGAAATCTATCGCCCAAGCTTCGGTTCGATAGCAGATAACCCAAATCAT 1423
QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
Db 1424 ATTGCACAAATTTTCATCGAGAGGAGCTACGAGGGATGGACGAATTAAGCCTGACGTAACA 1483
QY 221 AlaProGlyThrTyThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 1484 GCTCCTGGAAACATTTATTTTATCAGACCGCTTTCTTCTAGCTCCAGACTCTTCGTTTGG 1543
QY 241 AlaAsnHisAspSerLysTrpAlaTyMetGlyThrSerMetAlaThrProIleVal 260
Db 1544 GCGAATTTATAACAGTAAATACGCTATATGCGCGGTACCTCCATGGCGACACGCTATTGTT 1603
QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
Db 1604 GCAGGGAATGTCCGCAATTAGTGAGCATTTTATAAAAAATAGAGGTATTACTCTCTAG 1663
QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 1664 CTTCTTTAATAAAGCTGCACCTTATCGTGGTGTCTACTGATGTTGTTTAGGATATCCT 1723
QY 301 AsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320


```
Db 1724 AGTGTGACCAAGCGTGGCGGTGTTACTAGATAAATCGTTAAATAGCGGTATGTC 1783
QY 321 AaGluThrSerProLeuSerThrSerGlnLysAlaThrTySerPheThrAlaGlnAla 340
Db 1784 AATGAAGCAACTGATAGCCACAGACAAAAGCAAGTATTCGTTCCAGCACAGCG 1843
QY 341 GlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAlaSer 360
Db 1844 GGTAAACCTTTAAATAATCTCGTAGTAGGACAGATGCTCTGGAAGTCAACATGCATCT 1903
QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyVal 380
Db 1904 TATACACTAGTTAATGATTAGATCTAGTTATTACTGCTCCGATGACAAAATATGTA 1963
QY 381 GlyAsnAspPheThrAlaProTyAspAsnAsnTrpAspGlyAsnAsnValGluAsn 400
Db 1964 GGAATGATGTTTAGTTATTCCTTATGATAAATACTGGGATGTCGCAACAATGTTGGAAC 2023
QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyAsnVal 420
Db 2024 GTATTATAAAGCTCCGCAATCTGGAACGTATATAATGAGGTTCAAGCGTATAATGTA 2083
QY 421 ProValSerProGlnThrPheSerLeuLalleValHis 433
Db 2084 CCATCTGCCACACAGCGTTTCTCACTAGTATCTGATCAT 2122

RESULT 15
AF268611/c 60006 bp DNA linear BCT 30-MAY-2001
LOCUS
DEFINITION
Unclutered marine group II euryarchaeote EBAC37F11, BAC clone
37F11, partial sequence.
ACCESSION
AF268611
VERSION
GI:9664575
KEYWORDS
unclutered marine group II euryarchaeote 37F11
SOURCE
Archaea; Euryarchaeota; Marine Group II; environmental samples.
REFERENCE
1 (bases 1 to 60006)
Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
Construction and analysis of bacterial artificial chromosome
libraries from a marine microbial assemblage
Environ. Microbiol. 2 (5), 516-529 (2000)
21128653
JOURNAL
MEDLINE
PUBMED
11232180
AUTHORS
Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
Direct Submission
Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
FEATURES
location/Qualifiers
1..60006
/organism="unclutered marine group II euryarchaeote 37F11"
/mol_type="genomic DNA"
/db_xref="taxon:133822"
/cclone="BAC 37F11"
/note="environmental sample"
1484..3178
/note="37F11#1"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="S-layer protein"
/protein_id="AAF97179.1"
/db_xref="GI:9664576"
/translation="MEPQCTELIDIFVQSPSSAIAQVNIIVKRVSDGSCQGLEBEI
PVRVSPFENLNASQWFWSEGGYPLANNENTGNTMWSHTITLNLPGWNISGMQ
SASVSPHARGIPPEIIPPTNWDKQPLEIISIVHPQGMVSEITVHNSISFASTP
VLSSAGSSQAFSLHLSINPSTVTVPATMNODEYELILEEGINFETISSTNDIAIQ
```

CDS

```
/note="37F11#2"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97180.1"
/db_xref="GI:9664577"
/translation="MNQSNSTSAFYTLDCVSENLDTEFOALLEIRGNHQHWEKLV
NREAFRLKKQGLDEIDDRNKITWDDATLWFTCVRLNDRDSAPITLNGKNRFRGR
FPHTGSLQILENIIISQSEGLSDIEHLYLTERCSMTNGTKEFFLEGGKAGL
MLGYITFEVKALRSMLSGWVSEEPDGGVREAVRHLLNALLMAAERRGAGLIH
RQHR"
complement (3892..4797)
/note="37F11#3"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="SAICAR synthase"
/protein_id="AAF97181.1"
/db_xref="GI:9664578"
/translation="MGEMLYQGVKQVWSTDPPILFRFTNOISVFDQIIPSLIPRK
LEFNRYLTAHWFKLVEQGVGTHLVELNAPDRCLVRKVEVVGPAIPEDAEMVFPV
GESVRLVLSGSAWRFORGLTAELSIGQDAEYGVKLESEFVETTPKEADRID
REALQISNTIDQFDSIVQNALKYDEIIEERAAKGLHVDGKFKFALGPGKRVLV
DPTGDTDRWDAAEAYSNGECIELSKFEFVTRTHYINSHQELKLARDSGAIDPPIPA
LPESVIBETASLYSSMYERLTSQF"
complement (4846..5508)
/note="37F11#4"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="translation initiation factor 6"
/protein_id="AAF97182.1"
/db_xref="GI:9664579"
/translation="MPDITLDIHGSDQIGIHLSVGNVLFPHPRELPTPTIEKLEQVVG
LELAPIGINSNLITAGLNTKGMVADIESDIDALTAYGDVWVSGVNTAGNL
LVNEQGVASPSIPKQGLDILSELGVDIISTIGQDVGSLGVNSQGVLLHPDV
APEVLLIEBILGVPMVGVTFSGSPVVGAGICASNAGALAGSETTGPELNRDIALG
LI"
5597..5745
/product="5S ribosomal RNA"
5816..8611
/product="23S ribosomal RNA"
complement (8888..10192)
/note="37F11#7: contains Zn-ribbon and TPR-like repeats"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97183.1"
/db_xref="GI:9664580"
/translation="MNECAVCGILYLAGSPCAGSQLRQAOQSFSDSDAALPTEVPG
LDDAAQWYDLLEGIEAPVEQEEBAASLPFGGEGSTNLSRPLFGIGHGEGIPPEL
SNODGSESEVVOHVEPVIESNLPEVALETQHDSPVVSASVIEPVLAPEAPD
LPLIEEPVQVFPVGPRLTAIFPVEFTVDTSLPNEWSIESFGSDQIFTAEDVIE
VVFSDLEDTVVHVEHDGENTPQMEKOSVYSERLSLPDLHPAQLMVDVGTDSASRAL
LESQFTSIGQDMRASAARSFQIVARFPQDSGANNYNIGALLQVAVSWQHSNDFIYS
NSATQFEAILLSREAVTNSQSFSTYNLSQALLSGREERKALLDWMAGNSQRES
NFTNLRAAILAQGRYDEAKRLVLPQADKLASGNLILKLPAL"
complement (10248..11969)
/note="37F11#8"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="permease"
/protein_id="AAF97184.1"
```

rRNA

rRNA

CDS

CDS

```
/db_xref="GI:9664581"
/translation="MGNILFSSKRPQAQLIALISLLFVCTMOPISRSAAAGGDDGLD
TVIRDFENQAVYGVGILLFVYALITIEVHTLAAALGGLLAVIALNHSVREAL
SLKAVTMDNWTGMLLGVVWGIISHTGVPEFVAVQAYKSGSGEITWLVILCSV
TAVLSAPLDMETITMLLFTVIXIAQVLDLKPILISVLFISNIGGAATMGDPNI
MIGSAPDAISGNKPNIANLASDGVTFNDITEAPGLIMTIVFSLIRWLKDE
FSGRDRDITELERQAKIKGLKVSAILTILVILGFLHPVHIGVSKALKFFPDD
MLLATNRHELEPELVETWTLFLFAGLFLVHLSLOVMVDIVGYNVEKALKFFPDD
YLAALILILWVSATASADINIPYATMIPVLSLAPELNLPLNPLWALAFACIL
CGNGTLGASANVYAGMSEAGYPISEFPFRAGFPVLLSTFIVSFYMLIVYVGG
EDGALTWKIVLVGLISLLGIIAQYSRGRKGSFAELVDDDFDEIVSTIKSVVITGS
KGSEEE"
complement(12019..12282)
/notes="37F11#9; contains KH domain"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="RNA-binding protein"
/protein_id="AAF97188.1"
/db_xref="GI:9664583"
/translation="MSVFNHIIROAKDTSLPITMRIGKSLGSLTDSVVVELEGQLSSSL
VKVNRGLYPRDELKVTWNI MAEKTNSVVFARGNAVFEWRN"
complement(12279..12614)
/notes="37F11#10; contains Zn-finger; conserved protein"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97185.1"
/db_xref="GI:9664582"
/translation="MGRSRRLNRDKRGHALWSOTQLLSLLKNPSOFTPEQRSRFALH
LVKVRHRIRLPEVEIRMCRTCTILIRGENATIRFRNGHKIQTCHSCTSRRIPY
RQHSQKVKRA"
complement(12624..13274)
/notes="37F11#11"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="adenylate kinase"
/protein_id="AAF97187.1"
/db_xref="GI:9664584"
/translation="MDEGRQMLNIVLFGPPGAGKGTQQRMMDATGLPQVSTGDMVL
RAAVAGTSVGEIAKSYMDKALVPDSVIIIDRLDHDHDDAKNGMDFDGPRTVPOA
BALSLIVSVSAKLDVDPDIRIVERICRYSCTGTVYHDTFNPIQNGICGCSFTE
KRRADVEETVLARLEAVHTOTSLPADFYRAAGIFHVDGDRDIDVITDILLAVLG"
complement(13284..16691)
/notes="37F11#12"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97188.1"
/db_xref="GI:9664585"
/translation="WYMLQSGTSQATAVAGSASLARYLREVAGINKPSASLIKATL
INGAEDLTGPDIPNAGWQIDLENSLPSSGVSLSDFQDDERELQAGSLIYSD
LDGSKGIDITLTAASAEASANAQESRLNLDLILAPDGSYSLGNDFSSGISTTG
GADNLNIRIERIPAGATQNGDWMTVEHRGSSQSYIVIAADATLIPKADLTTF
GNSILPSSSTPLIGLVTSLAHNOGLASGPYRIQEDITSGSILYDSNRSLLEGG
SLDSVSTFQCTTGVHLRLSLDNNQVPELNDAINDNIIIDLEITTAQGLRVI
FONQGSPTTQSDRDSASVMDVNRNGLSLPVIAREGTERGVSLTVSSVQEPD
PYPLILUSPEDSKSVNQTGVFTISGQEDNFTIYLALDSDTSFAGATKRYAR
AGSFIVDVTRYQNTPTSHQRYLIEIGQVDSVDVPSGTSVSAKFGERTGYSIGV
RNTGNSAQYTCVSTSSQWQIMLGNSSLSLEFPLNLQDLSMDNVNFIPISTANG
PLAGSTDOVCTVGTDSLNRVTVDAVSELKAFRSLDYPDGPAGPAGALPVF
ADTGLVYFNHTIQNGNVPDPAVTLERNGPMAEIQFDEQISSTLSVTLPGQS
GVNRELLVPEIRNGNSNTYLRVESSPQMTLNSILVNGNLVSNVGTLIA
APVNDFTTEIRVENSGLDLDLEWSTSLADPCHSICYSNPTSPVPSVSAQVLA
KAPNQTASGFQDMQLFVNGNNGNPTNABLLRVLEAATSFAGLSDEFTIALLFI
PRDGSQKQITITNENITPLSQLTVEVRDGNLTDRSTISSEITDLAIGESIE
VGVKTSQSDSDIDGRMLVILITDGVIFEMADTYSVSSQSGGIFGILPAYSIP
LVLLALLGVYGRRLKQSKMDDGTVELVAPDAHTDADHLGTRREQALDISVNDI
ASGEVSDQIEAALAAQSLDMPVPTKKAQVPTGRPPSGLPSMGLPVGLPAGLPAGL
```

```
CDS
PPAKSVFLPVQVTAGPPLPPGGLPDGWTMEQNNHYGEQYLQRMGLN"
complement(16790..18130)
/notes="37F11#13"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="membrane-associated subtilysin-type serine
protease"
/protein_id="AAF97189.1"
/db_xref="GI:9664586"
Alignment Scores:
Pred. No.: 6,838-29 Length: 60006
Score: 626.50 Matches: 167
Percent Similarity: 53.52% Conservative: 76
Best Local Similarity: 36.78% Mismatches: 156
Query Match: 27.82% Indels: 56
DB: 1 Gaps: 17
US-09-985-689a-7 (1-433) x AF268611 (1-60006)
QY 13 AlaGlnAsnAsnPheGlyLeuTyrglyGlnGlyGlnIleValAlaValAlaAspThrGly 32
Db 17383 GTCCTCTCTCTTGAATCTTGATGTTCTGGAGAAGCATTCGATGATGATCGC 17384
QY 33 LeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAla 52
Db 17323 GTCGATATGGAT-----CATCTGCCTCATCGAGGGTTGCAGCA 17282
QY 53 LeuTyrAla---LeuGlyArgThrAsnAsnAlaAsnAspProAsn---GlyHisGlyThr 70
Db 17281 ATTAATACTCAGTTCGGTTTATAGATAGTTCTCCAAATGATAGCAACAGTGTGATGGAGC 17222
QY 71 HisValAlaGlySerValLeuGlyAsnAlaThrAsn-----LysGlyMetAla 86
Db 17221 CAGTGGCTATGACTGTCTTGGCATGTTACGGGTATGATGATGATGATGATGATGATGATGAT 17162
QY 87 ProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeu 106
Db 17161 CCAGNAGCAATCTTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 17102
QY 107 ProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrsrAlaGlyAlaArgIleHisThr 126
Db 17101 CAAGGTTCATATCATGATCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 17042
QY 127 AsnSerTyrGlyAlaProValAsn---GlyAlaTyrsrThrAspSerArgAsnValAsp 145
Db 17041 AATGATGGGTCTTAATGGAATTTACGAGGCTACACTGCTGATCA-AGATCGACAGAT 16983
QY 146 AspTyrValArgLysAsnAspMetThrIle---LeuPheAlaAlaGly---AsnGluGly 163
Db 16982 CAATATGCTCTCCAAATAATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 16923
QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 16922 ACAGGTCTTCAAGATTAATCTCTCCAGNACAGCAAGATGTTCTTCCATGCTCTCA 16863
QY 184 ThrGluAsnLeuArgProSerPheGlySerTyrsrAlaAspAsnIleAsnHisValAlaGln 203
Db 16862 AGCGTTAAAT-----GGTAGC-----GTCAGCTCC 16839
QY 204 PheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaProGly 223
Db 16838 TATTTCATCGGAAGTCTCACTCTCGATGGAGNATTAACCTGATTATTAGTCTCGCTCCGGT 16779
QY 224 ThrTyrlsLeuSerAlaArgSerSerLeuAlaProAsp----- 236
Db 16778 GACGGTATCTGTTCTGCAGCAGCAGGAGGAAGCAAGCAAGCTTGTGTTGTTGTTGTTGTTGTTGTT 16719
QY 237 SerSerPheThrAlaAsnHisAspSerLysTyrsrAlaTyrsrMetGlyGlyThrSerMetAla 256
Db 16718 ATTCGTTACTCACTCAATATGTTGTTCAATGATATATGCAACTCAGCGGAACCTTCTCAGCA 16659
```

[illegible]

Search completed: April 4, 2004, 08:10:14
Job time : 3595.27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 3, 2004, 23:50:44 ; Search time 63.4512 Seconds
(without alignments)
3787.066 Million cell updates/sec

Title: US-09-985-689A-7
Perfect score: 2252
Sequence: 1 NDVARGIVKADVAQNFGLY.....EVQAVNPVSPQTFSLAIVH 433

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US0985689/runat_31032004.161807.4184/app_query.fasta_1.3498
-DB=Issued Patents NA -Qfmt=fastp -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0985689 @C@N 1.1 142 @runat_31032004.161807.4184 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Issued Patents NA:*
- 2: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2252	100.0	3003	2	US-08-873-479-41
2	2125.5	94.4	1923	4	US-09-509-814A-5
3	2120.5	94.2	1923	4	US-09-509-814A-7
4	2110.5	93.7	1920	4	US-09-509-814A-3
5	432.5	19.2	1977	3	US-08-894-818B-2
6	432.5	19.2	1977	4	US-09-445-472-11
7	396	17.6	1236	4	US-08-445-472-2
8	396	17.6	1566	3	US-08-894-818B-4
9	396	17.6	1962	3	US-08-894-818B-34
10	396	17.6	1962	4	US-09-445-472-15
11	390	17.3	1977	3	US-08-894-818B-6
12	332.5	14.8	2539	3	US-09-000-016-3

13	332.5	14.8	2539	4	US-09-514-340-3	Sequence 3, Appli
14	327.5	14.5	2809	3	US-09-000-016-1	Sequence 1, Appli
15	327.5	14.5	2809	4	US-09-514-340-1	Sequence 1, Appli
16	311.5	13.8	2835	1	US-08-750-532-2	Sequence 2, Appli
17	311.5	13.8	4765	1	US-08-750-532-8	Sequence 7, Appli
18	311.5	13.8	4765	3	US-08-894-818B-7	Sequence 8, Appli
19	311.5	13.8	4765	4	US-09-445-472-5	Sequence 5, Appli
20	309	13.7	2532	1	US-07-671-376C-4	Sequence 4, Appli
21	297	13.2	1859	3	US-08-894-818B-15	Sequence 15, Appli
22	276	12.3	898	1	US-08-750-532-7	Sequence 7, Appli
23	271.5	12.1	1306	4	US-09-966-921A-1	Sequence 1, Appli
24	271.5	12.1	1330	4	US-09-966-921A-5	Sequence 5, Appli
25	270	12.0	564	1	US-08-750-532-11	Sequence 11, Appli
26	270	12.0	564	3	US-08-894-818B-14	Sequence 14, Appli
27	257.5	11.4	807	1	US-07-706-691G-56	Sequence 56, Appli
28	257.5	11.4	807	2	US-08-354-021-56	Sequence 56, Appli
29	257.5	11.4	807	3	US-08-618-446-56	Sequence 56, Appli
30	257.5	11.4	807	3	US-08-980-135-56	Sequence 56, Appli
31	257.5	11.4	807	4	US-09-585-798-56	Sequence 56, Appli
32	256.5	11.4	807	1	US-08-566-369-2	Sequence 2, Appli
33	256.5	11.4	807	3	US-09-074-331-2	Sequence 2, Appli
34	256.5	11.4	807	5	US-09-074-331-2	Sequence 2, Appli
35	255.5	11.3	807	1	US-07-706-691G-57	Sequence 57, Appli
36	255.5	11.3	807	1	US-08-254-021-57	Sequence 57, Appli
37	255.5	11.3	807	2	US-08-618-446-57	Sequence 57, Appli
38	255.5	11.3	807	3	US-08-980-135-57	Sequence 57, Appli
39	255.5	11.3	807	4	US-09-585-798-57	Sequence 57, Appli
40	253.5	11.3	807	1	US-08-566-369-5	Sequence 5, Appli
41	253.5	11.3	807	3	US-09-074-331-5	Sequence 5, Appli
42	253.5	11.3	807	5	US-09-074-331-5	Sequence 5, Appli
43	253	11.2	10216	2	US-08-875-154-1	Sequence 1, Appli
44	251.5	11.2	807	1	US-08-566-369-8	Sequence 8, Appli
45	251.5	11.2	807	3	US-09-074-331-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-08-873-479-41

; Sequence 41, Application US/08873479

; Patent No. 5891701

; GENERAL INFORMATION:

; APPLICANT: Sloma, Alan

; APPLICANT: Lyne, Christanson

; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide

; TITLE OF INVENTION: Having Protease Activity

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 58917010 No. 5891701disk of No. 5891701th America

; STREET: 405 Lexington Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/873,479

; FILING DATE: 12-JUN-1997

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Agri, Cheryl H

; REGISTRATION NUMBER: 34,086

; REFERENCE/DOCKET NUMBER: 5251.000-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-857-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 41:

; SEQUENCE CHARACTERISTICS:

41699 102b

```
; LENGTH: 3003 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 1,86e-222 Length: 3003
Score: 2252.00 Matches: 433
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-09-985-689a-7 (1-433) x US-08-873-479-41 (1-3003)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPhedGlyLeuTyr 20
Db 1470 AATGACGTGGCCGTGGCATTGTGAAGCAGACGTCGCACAAATAACTTTGGCTTATAT 1529

QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGACAGGACAGATTGATGAGGATGCTGATACCTGGGCTTGATACAGGAGAAATGACAT 1589

QY 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGATGAAGCATTCGCGGTGAAGATTACCGCACTATATGCACTGGCGCAGACGAAT 1649

QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnAla 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGATCTGTTAGGAATGCT 1709

QY 81 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 100
Db 1710 ACAATTAAGGATGGCAGCCGCAACCAATCTAGTCTTCAATCTATATGATAGTAGGT 1769

QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrIlePheSerGlnAlaTyrSerAla 120
Db 1770 GGAGGGCTGGGAGGACTACCTGCTAATCTCAAAATTAATTCAGTCAAGCAATATAGTCT 1829

QY 121 GlyAlaArgIleHisThrAsnSerTyrGlyAlaProValAsnGlyAlaTyrThrAsp 140
Db 1830 GGAGGCGAATTCATACGAATTCATGGGGCTCCAGTAAACGGTGCCTATACGACAGAC 1889

QY 141 SerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGly 160
Db 1890 TCTCGAATGTGTGATGATTATGTGAGAAAAATGATATGACGATCTTTTTGCGGCCGGA 1949

QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 1950 AATGAGGACCCAGGTAGCGGTACATCAATCAATGTCACCGGACAGCAAAAAATCGGATTACA 2009

QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHis 200
Db 2010 GTTGGGCAACCGAAACCCAGTCCGATCTTATGCGGATAAATATTAAACAT 2069

QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyValGlyIleLysProAspValMet 220
Db 2070 GTTGCTCAATCTCTTCCAGAGGTCTCTAGAGATGGAGCTATTAAACCGGACGTCATG 2129

QY 221 AlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTyr 240
Db 2130 GCACCAAGGTACGTATATCTCTGCTAGATCATCATTAGCTCCAGATTCTCTCTCTGG 2189

QY 241 AlaAsnHisAspSerIleTyrAlaTyrMetGlyThrSerMetAlaThrProIleVal 260
Db 2190 GCAACCATCATGATAGTAATATGCTACATGGGGTGGTACTCTATGCTACTCTCAATGTA 2249

QY 261 AlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys 280
Db 2250 GCAGGTAAATGTTCCACAAATTAAGGGAGCATTTTGTGAAAAATAGAGGGGTAACTCTTAAG 2309
```

```
QY 281 ProSerLeuLeuLysAlaAlaLeuLysAlaGlyAlaAlaAspValGlyLeuGlyPhePro 300
Db 2310 CCTTCCCTTTTAAAGCTGCTTTAAATTCAGAGTCTGCGGATGTGGACTTGGCTTTCCA 2369

QY 301 AsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPheVal 320
Db 2370 AATGTAACCAAGGATGGGAGAGTAACGTTAGATTAATCCCTTAATGTCGCAATTTGTG 2429

QY 321 AsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAla 340
Db 2430 AATGAAACGAGCCCTTTATCAACAAAGTCAAAAAGCAACATATTCGTTACGGCTCAAGCT 2489

QY 341 GlyLysProLeuLysIleSerLeuValTyrSerAspAlaProGlySerThrThrAlaSer 360
Db 2490 GGTAACCCCTTAAATAATATCATCTTTTGGTCAGATGCAACGAGTAGCAGCGCATCA 2549

QY 361 LeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrVal 380
Db 2550 CTAACCTTTAGTAATGATTAGATTAGCTTACCTACCTGCAACCAATGGAACCTAAATACGTC 2609

QY 381 GlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGluAsn 400
Db 2610 GGAATGACTTTACAGCACCGTATGATAACAAATTTGGGATGGCAGAAACACAGCTGGAAT 2669

QY 401 ValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyrAsnVal 420
Db 2670 GTGTTTATCAATGCTCTCAAGCGGACGATATACAGTCGAAGTCGAGGCTTACAATGTA 2729

QY 421 ProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db 2730 CCAGTAAGTCCGCAACCTTTTCTTTAGCGATTGTACAT 2768
```

```
RESULT 2
US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Pred. No.: 1,07e-209 Length: 1923
Score: 2125.50 Matches: 406
Percent Similarity: 97.93% Conservatives: 19
Best Local Similarity: 93.55% Mismatches: 8
Query Match: 94.38% Indels: 1
DB: 4 Gaps: 1

US-09-985-689a-7 (1-433) x US-09-509-814A-5 (1-1923)
```



```
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db : : : : :
859 TCACATATATAAGGAATGGCGCTCAGCGCAATCTAGTCTTCCAATCTATCATGGATAGC 918
QY 100 GlyGlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaIleVal 119
Db : : : : :
919 GGTGGGGGACTTGGAGGACTACCTTCGAATCTCGAAACCTTATTCAGCCAAAGCATACAGT 978
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaProValAsnGlyAlaIleVal 139
Db : : : : :
979 CTGGTCCAGAAATCTATCAAACTCTCGGGAGCAGCACTGAATGGGGCTTACACAACA 1038
QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
Db : : : : :
1039 GATTCCAGAAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTCTTTTCGGTGCC 1098
QY 160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
Db : : : : :
1099 GGGAAATGAGGACCGAAGCGGGGAAACCATCAGTGCACAGGCACAGCTAAATGCAATA 1158
QY 180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 199
Db : : : : :
1159 ACAGTCGGAGCTACGGAAACCTCGGCCAAGCTTTGGGTCTTATGCGGCAATATCAAC 1218
QY 200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
Db : : : : :
1219 CATGTGGCACAGTCTCTTCACGTGGACCGCAAGGATGACGAGTCAACCGGATGTC 1278
QY 220 MetAlaProGlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
Db : : : : :
1279 ATGGCACCGGGAAGCTTCTACTACTATCAGCAAGATCTTCTCTTGACCGGATCTCTCCTC 1338
QY 240 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 259
Db : : : : :
1339 TGGCGAACCATGACAGTAATATGCAATACATGGGTGGAGCTCCATGGCTTACACCGATC 1398
QY 260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
Db : : : : :
1399 GTTGTGGAAAGCTGGCACAGCTCTGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
QY 280 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db : : : : :
1459 AAGCCTTCTCTATTAAGAGCGCACTGATTCGGGTGAGCTGATCATCGGCTTGGGTAC 1518
QY 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db : : : : :
1519 CCGAACGTTAACCAAGATGGGACGAGTGCATTTGGATTAATCCCTGAACGTTGCCTAT 1578
QY 320 ValAsnGluThrSerProLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaGln 339
Db : : : : :
1579 GTGAACGAGTCCAGTCTCTATCCACCAAGCAAGCAAGGAGCGATCTCTCTTACTGCTACT 1638
QY 340 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProGlySerThrAla 359
Db : : : : :
1639 GCCGGACAGCTTTGAAATACTCTCTGGTATGGTCTGATGCCCTCGAGGACACACTGCT 1698
QY 360 SerLeuThrLeuValAsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyr 379
Db : : : : :
1699 TCCGTAACGCTTGTCAATGATCTGACCTTGTCAATACCGCTTCCAAATGGCAGACAGTAT 1758
QY 380 ValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGlu 399
Db : : : : :
1759 GTAGGAATGACTTCTTCCGCAATACATGATACTGGGATGGCGGCAATTAACGTAGAA 1818
QY 400 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaIleVal 419
Db : : : : :
1819 AATGTATTATTAAATGACCAACAAAGCGGACGATATCAATTAAGTACAGGCTTATAAC 1878
QY 420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
Db : : : : :
1879 GTACCGGTTGGACCAAGCAACTCTCTGTTGGCAATTTGTGAAT 1920
```

RESULT 4

```
US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-OPCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3
```

Alignment Scores:

Pred. No.:	3,79e-208	Length:	1920
Score:	2110.50	Matches:	404
Percent Similarity:	97.24%	Conservative:	18
Best Local Similarity:	93.03%	Mismatches:	11
Query Match:	93.72%	Indels:	1
DB:	4	Gaps:	1

US-09-985-689A-7 (1-433) x US-09-509-814A-3 (1-1920)

```
QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTyr 20
Db 616 AATGATGTGGCAGAGGATATTGTCAAGCGGATGTGGACAGAGGACTACGTTTGTAT 675
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGAACAAGCCAGATTGTGCGAGTTGCGGATCTGCGATGATGATACAGGAACACGACAT 735
QY 41 SerMetHisGlnAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCGATGAAGCCCTCCGCGTAAATACAGCAGCTATATGCACTGGGTGGAGCAAT 795
QY 61 AsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn--- 79
Db 796 AATGCGAATGATACGAACCGTTCATGTTACCCATGTCGAGGTTCGGTATTAGGAATGCG 855
QY 80 AlaThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 99
Db 856 GCAACGAATAAAGAAATGGCACCTCAAGCGAATCTGGTTTTTCAATCCATCATGGATAGC 915
QY 100 GlyGlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 119
Db 916 AGTGGTGGGCTTGGAGGCTTGCCTTCCAATCTGCAAACTTATTTCAGCAAGCATTCAGT 975
QY 120 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThr 139
Db 976 GCAAGGTGCCAGAAATTCATCAAACTCTCTGGGGGCGAGCGGTGAATGGGGCTTACAGCA 1035
QY 140 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 159
Db 1036 GATTCCAGAAATGTGGATGACTATGTAAGGAATAATGATATGACGATCTCTTTCGGGCT 1095
```

160 GlyAsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 179
1096 GGGATGAAGAGCGGACCGGATGACCTGACCTGATCGGCTAAAGCCGATA 1155
180 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAsn 199
1156 ACAGTCGGCGCAACCGAAACCTCGCTCAACCTTCGCTGATGAGATTAATTAAC 1215
200 HisValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspVal 219
1216 CAGTTTGACAGTTCCTTCCTCCGCGCGACAAAGATGGCGAATCAAGCTGATGTC 1275
220 MetAlaProGlyThrThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 239
1276 ATGGCCCGGAGGACATACATATTTATCAGCAAGATCTTCTGTGACCGGATTCCTCTTC 1335
240 TrpAlaAsnHisAspSerLysThrAlaThrMetGlyThrSerMetAlaThrProIle 259
1336 TGGCGGAATCATGACAGCAATATGCTATATGGTGGACGCTCCATGGCAACCGGATT 1395
260 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
1396 GTTCCGGGGAATTTCCAGACGCTCCGTGAGCATTTTGTAATAATAGAGGAATCACTCT 1455
280 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspValGlyLysPhe 299
1456 AAGCTTCCTATTGAAGCAGCTTGAATGTCAGCTGCTGATGTTGGATTGGGTTAT 1515
300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
1516 CCGAAGGAAACCAAGATGGGCGGAGTGACCTGGGATAATCGTTGAAGTTGGCTAT 1575
320 ValAsnGlyThrSerProLeuSerThrSerGlnLysAlaThrThrSerPheThrAlaGln 339
1576 GTGAAGCAATCCAGTCCCTATCAACTAGCAAAAGGACATATACCTTTACTGCAACG 1635
340 AlaGlyValProLeuLysIleSerLeuValThrSerAspAlaProGlySerThrThrAla 359
1636 GCGGCGAGGCAATGAAATCTCCCTGGTATGTCGGATGCCCTGCAAGCACTACTGCT 1695
360 SerLeuThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrLysThr 379
1696 TCTGTAACTCGTCAATGATTGGATTGGTCAATCAGACCAACCAAGCAAGATAT 1755
380 ValGlyAsnAspPheThrAlaProThrAspAsnAsnTrpAspGlyArgAsnValGlu 399
1756 GTCCGGAATGACTTTCAGACCAATTTGCAATATCTGGATGGGATGCGCCGCAATACCTGAA 1815
400 AsnValPheIleAsnAlaProGlnSerGlyThrThrThrValGluValGlnAlaThrAsn 419
1816 AATGTATTATTAAATTCGCCCAAGTGGAAACATATACCATTTGAGTGTCAAGCATATAAT 1875
420 ValProValSerProGlnThrPheSerLeuAlaIleValHis 433
1876 GTCCCGGTGGACCAACAACTCTCTGTTGGCAATTGTGAAC 1917

RESULT 5

US-08-894-818B-2

Sequence 2, Application US/08894818B

Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru

APPLICANT: MORISHITA, Mio

APPLICANT: YAMAMOTO, Katsuhiko

APPLICANT: MITTA, Masanori

APPLICANT: ASADA, Kiyozo

APPLICANT: TSUNAWA, Susumu

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENES

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:

Pred. No.: 8,91e-35 Length: 1977
Score: 432.50 Matches: 134
Percent Similarity: 44.85% Conservative: 62
Best Local Similarity: 30.66% Mismatches: 156
Query Match: 19.21% Indels: 85
DB: 3 Gaps: 16

US-09-985-689A-7 (1-433) X US-08-894-818B-2 (1-1977)

QY 8 ValLysAlaAspValAlaGlnAsnAsnPheGlyLeuTy-GlyGlnGlyGlnIleValAla 27
Db 433 ATAGGGCGGATACCGCTCGTGAACCTCCCTCGGTACGACGGAAGCGGTGTGGTGTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCTGTCGATACGGGTATAGACGGGAAAC-----CACCCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GCGAAGGTCATAGCTGCTGACGCGCTCAACGGCAGGTGACCCCTACGATGACCGAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82
Db 595 GGACACGGAACCCACGTTGGGGGTATCTGTCGCCGAACCGGACGGCTTAACCTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100
Db 655 ATAGCGTCCGCCCGCGGGAAGCTCTGTCGGGTCAAGGTCTCGGTCCGACCGGTTCG 714
QY 101 GlyGlyLeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 120
Db 715 GGAAGCGTCTCCACCATCATCTGGGTGTGACTGGGTGCTCCAGAACCAAGGACAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyValaProValaAsnGlyAlaTyrThrAsp 140
Db 775 GGGATAGGGTCATCAACCTCTCCCTCGGCTCTCCCGAGAGCTCCGACGGAACCGACTCC 834


```

; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

Alignment Scores:
Pred. No.: 8,91e-35 Length: 1977
Score: 432.50 Matches: 134
Percent Similarity: 44.85% Conservative: 62
Best Local Similarity: 30.66% Mismatches: 156
Query Match: 19.21% Indels: 85
DB: 4 Gaps: 16

US-09-985-689A-7 (1-433) x US-09-445-472-11 (1-1977)
QY 8 ValLysAlaAspValAlaGlnAsnPhelGlyLeuTyGlyGlnGlyValAla 27
Db 433 ATAGGGCGGATACCGTCTGGAACTCCCTCGGTACGACGAGCGGTGTGGTGGC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisGluAlaPheArg 47
Db 493 ATCTCGATACGGGTATAGACCGGAAC-----CACCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTy---AlaLeuGlyArgThrAsnAsnAlaAsnAspProAsn 66
Db 535 GGCAAGGTATAGGTGGTACACGCGGTCAACGGCAGGTGACCCCTACGATGACCAAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsn 82
Db 595 GGACACGGAAACCCACGCTTCGGGTATCGTTGCCGGAACCGGACGGGTAACTCCAGTAC 654
QY 83 LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 100
Db 655 ATAGGGCGTCCCGCGCGGCGGAGCTCGTGGCGTCAAGGTTCTCGGTCCGACCGGTTCG 714
QY 101 GlyGlyLeuGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 120
Db 715 GGAAGCGTCTCCACCATCATCGGGGTGTGACTGGTCTCCAGAACCAAGCAAGTAC 774
QY 121 GlyAlaArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrAsp 140
Db 775 GGGATAAGGTCTATCAACCTCTCCCTCGGTCTCTCCAGAGCTCCGACCGAACCGACTCC 834
QY 141 SerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaGly 160
Db 835 CTAGTCAGCGCTCAACACGCGCTGGACCGCGGTATAGTAGTCTCGTCCGCGCGCGC 894
QY 161 AsnGluGlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThr 180
Db 895 AACACGGCGGCGGAACACCTACACCGTCTCGGTCTCCCGCGCGCGCGAGGTCATAACC 954
QY 181 ValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHis 200
Db 955 GTCGGTGCA-----GTTGACGACCAAGCAAC 981
QY 201 ValAlaGlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMet 220
Db 982 ATCCGACGTTCTCCAGCAGGGACCGACCGCGGAGGTCAGCGGAGGTCAAGCGGAAGTCGTC 1041
QY 221 AlaProGlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrp 240
Db 1042 GCCCGCGGCTTGACATCATAGCCCGCGCGCGCAGC-----GGAACACGATGGGC 1092
QY 241 AlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIleVal 260
Db 1093 ACCCGGATAACGACTACTACACGAGCCCTCTGGAACGATGGCCACCCCGCAGCTT 1152
QY 261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
Db 1153 TCGGCGTGGCGGCTCATCTCCAGGCCAC-----CCG 1188
QY 280 Lys-----ProSerLeuLeuLysAlaAlaIleAlaGlyAla-----292
Db 1189 AGCTGGACCCCGGCAAGGTGAACCGCCCTCATCGAGCCGCGACATGTCGCCCC 1248
QY 293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgVal 309
Db 1249 AAGGAGATAGGGACATCGCTACGGTCCG-----GGTAGGGTG 1287
QY 310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326
Db 1288 AACGCTCAAGGCGCATCAAGTACGACGACTACGCCAAGCTCACTTACCGGCTCCGTC 1347
QY 327 SerThrSerGlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysIle 346
Db 1348 GCCGACAGGGAGCGCCACCCACCTTC-----1377
QY 347 SerLeuValTrpSerAspAlaProGlySerThrAlaSerLeuThrLeuVal-----364
Db 1378 -----GAGCTCAGCGCGCCACCTTCGTGACCGCCACCCCTCTACTGGAC 1422
QY 365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyValGly 381
Db 1423 ACGGGCTCGAGGACATCGACCTCTACCTTACGACCCCGACCGGAACGAG---GTTGAC 1479
QY 382 AsnAspPheThrAlaProTyTrpAspAsnAsnTrpAspGlyArgAsnValGluAsnVal 401
Db 1480 TACTCTACACCGCTTACTAC-----GGCTTCGAGAAGTTC 1515
QY 402 PheIleAsnAlaProGlnSerGlyThrTyThrValGluValGlnAlaTyR 418
Db 1516 GGCTACTACACCCGACCGCGGACCTTGAGCGGTCAAGGTCTGACGTAC 1566

RESULT 6
US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, HiKaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6

```

1093 ACCCGATAAAGCACTACTACACCAAGCGCTCTGGAACCAAGCATGGCCACCGCACGTT 1152
261 AlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrPro 279
1153 TCGGGCGTGGCGGCTCATCTCCAGGCCAC-----CCG 1188
280 Lys-----ProSerLeuLeuLysAlaAlaLeuLleAlaGlyAla-----292
1189 AGCTGGACCCCGGCAAGAGTGAAGCCGCTCTCATCGAGACCGCCGACATAGTCGCCGCC 1248
293 -----AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTTPGlyArgVal 309
1249 AAGGAGATAGCGGACATCGCTACGCTGCG-----GATAGGGTG 1287
310 ThrLeuAspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeu 326
1288 AACGTCTACAAGGCGCATCAAGTACGAGCACTACGCCCAAGCTCACCTTACCGGCTCCGTC 1347
327 SerThrSerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIle 346
1348 GCGGACAAAGGAAGCGCCACCCACCTTC-----1377
347 SerLeuValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal-----364
1378 -----GAGCTCAGCGCGGCCACCTTCGTGACCGCCACCTCTACTCGGAC 1422
365 -----AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGly 381
1423 ACGGGCTCGAGGACATCGACTCTACCTCTACGACCCCAAGCGGAAGAG-----GTTGAC 1479
382 AsnAspPheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnValGluAsnVal 401
1480 TACTCTACACCGCTACTAC-----GGCTTCGAGAAGGTC 1515
402 PheIleAsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyr 418
1516 GGCTACTACACCGCGCGGCACTGAGCGTCAAGTCTGCTCAGCTAC 1566

RESULT 7

US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOTO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: 1999-12-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2

Alignment Scores:

Pred. No.:	2,56-31	Length:	1236
Score:	396.00	Matches:	127
Percent Similarity:	42.56%	Conservative:	56
Best Local Similarity:	29.53%	Mismatches:	163
Query Match:	17.58%	Indels:	84
DB:	4	Gaps:	15

US-09-985-689A-7 (1-433) x US-09-445-472-2 (1-1236)

16 AsnPheGlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
58 AACTTGGATATGATGTTCTGGAATCAACATAGGAATTAATGACACTGGATATGAC---114
36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
115 -----GCTTCTCATCCAGATCTCCAGGAAAGTA-----144
56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
145 ATTGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGA 204
70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
205 ACTCATCTAGCTTCAATAGCAGCTGGTACTGGAGCAAGTAATGGCAAGTACAGGGA 264
85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGly 102
265 ATGGCTCAGGAGTAAGCTGGCGGGAATTAAGTTCTAGTCCGATCGGTCTGGAAGC 324
103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
325 ATATCTACTATAATTAAGGAGTTGAGTGGCGCTTGATAACAAAGATAAGTACGGAAT 384
123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyValaTyrThrThrAspSerArg 142
385 AAGTCTAATCTTCTCTGTTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 444
143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
445 CAGGCTCTTAATCAGCGTGGGATGCTGGATTAAGTTGTTGTTGCTCCGCTGGAAACGT 504
163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
505 GGACCTAACAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTTATTACAGTTGA 564
183 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202
565 GCC-----GTTGACAAGTATGATGTTATAACA 591
203 GlnPheSerSerArgGlyProThrArgAspGlyValIleLysProAspValMetAlaPro 222
592 AGCTTCTCAAGCAGAGGCGCAACTCGACAGCGCAGCTTAAGCCTGAGTTCTGCTCCA 651
223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
652 GGAACCTGGATAATTCCTGCCAGAGCAAGT-----GGAACTAGCATGGTCAACCA 702
243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
703 ATTAATGACTATTACACAGCAGCTCTCGGACATCAATGGCAACTCTCAGCTAGCTGCT 762
263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys-----280
763 ATTGCAGCCCTCTTGTCTCCAA-----GCACACCCGAGCTGGACT 801
281 ProSerLeuLeuLysAlaAlaLeuLleAlaGlyAla-----292
802 CCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAAATA 861
293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAsp 312
862 GCGGATATAGCTACGTCGCA-----GGTAGGGTTAATGCATAC 900
313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329
901 AAGCTATAAAGTACGATAAATGCAAGCTAGTGTCTCAGTGGATATGTTGCCAACAAA 960
330 GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349
961 GGCAGCCAAACTCACCAGTTCTGTTATTAGCGGAGCTTCGTTAAGTAACTCCCATATATAC 1020

QY 350 TipSexAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
Db 1021 TGGACAATGCCAAT-----AGGACCTTGACTT 1050
QY 370 ValIleThrAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 389
Db 1051 TACCTCTAGATCCCAATGGAACACAG--GTGACTACTCTTACACCGCTACTAT-- 1104
QY 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
Db 1105 -----GGATTCGAAAGGTTGTTATTACAAACCACTGATGGA 1143
QY 410 ThrTyThrValGluValGlnAlaTyAsn 419
Db 1144 ACATGGACAATTAAAGGTTGTAAGCTACAGC 1173

RESULT 8

US-08-894-818B-4
Sequence 4, Application US/08894818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
OTHER INFORMATION: /note= N at position 1283 is G or T.

US-08-894-818B-4

Alignment Scores: 3.63e-31 Length: 1566

Pred. No.:

Score: 396.00 Matches: 127
Percent Similarity: 42.56% Conservative: 56
Best Local Similarity: 29.53% Mismatches: 163
Query Match: 17.88% Indels: 84
DB: 3 Gaps: 15

US-09-985-689A-7 (1-433) x US-08-894-818B-4 (1-1566)

QY 16 AsnPheGlyLeuTyGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
Db 58 AACTTGGGATATGATGTTCTGGAATCAATAGGAATAATTGACACTGCAATTGAC-- 114
QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyIleThrAlaLeuTyAla 55
Db 115 -----GCTTCTCATCCAGATCTCCAGGAAGAATA----- 144
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 145 ATTGGGTGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGACATGGA 204
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLeuGly 84
Db 205 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGACGACCAAGTAATGGCAAGTACAGGA 264
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGly 102
Db 265 ATGGCTCCAGGAGCTTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGTTCTTGAAGC 324
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAla 122
Db 325 ATATCTACTATAATTAAGGGAGTTGAGTGGCGCTGATTAACAAGTAAGTACGGAATT 384
QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSerArg 142
Db 385 AAGGTCAATTAATCTTCTCTTGGTTCAAGCCAGAGCTCAGATGCTACTGACGCTCTAAGT 444
QY 143 AsnValAspAspTyrrValArgLyAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
Db 445 CAGGCTGTAAATGACGCTGGATGCTGGATTAGTTGTTGCTGCGCTGGAACAGT 504
QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 505 GGACCTAAACAAGTATACAACTCGTTCTCCAGCAGCTGCAAGCAAAAGTTATTACAGTTGA 564
QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyrrAlaAspAsnIleAsnHisValAla 202
Db 565 GCC-----GTTGACAGTATGATGTATAACA 591
QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
Db 592 AGCTTCTCAAGCAGAGGCGCAACTGCAGCGCAGGCTTAAGCTGAGGTTGTTGCTCCA 651
QY 223 GlyThrTyrrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 652 GGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCATGGGTCAACCA 702
QY 243 HisAspSerLysTyrrAlaTyrrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 703 ATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCAACTCTCTCAGTAGCTGGT 762
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
Db 763 ATTGACGCGCTCTTGTCTCAA-----GCACACCGAGCTGGACT 801
QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
Db 802 CCAGACAAAGTAAAAACAGCCCTCATAGAACTCTGATATCGTAAAGCCAGATGAATA 861
QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgGlyValThrLeuAsp 312
Db 862 GCCGATATAGCTACGGTGCA-----GGTAGGTTAATGCATAC 900
QY 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329

Db 901 AAGGCTATAAAGTACAGTAAGTATCAAGTATGTTTCAGTATGTTGCGCAACAAA 960
QY 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349
Db 961 GCGAGCCAACTCACCAGTTCGTTATTAGCGAGCTTCGTTGTAAGTCCACATATATAC 1020
QY 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
Db 1021 TGGGCAATGCCAAT-----AGCGACCTTGATCTT 1050
QY 370 ValIleThrAlaProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyAsp 389
Db 1051 TACCTCTACGATCCCAATGAACAG--GTTCAGTACTCTTACACCGCTACTAT-- 1104
QY 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
Db 1105 -----GGATTCGAAAGGTTGGTTATTACAACCCCAACTGATGGA 1143
QY 410 ThrTyThrValGluValGlnAlaTyAsn 419
Db 1144 ACATGACAAATTAAGTTGTAAGCTACAGC 1173

RESULT 9

US-08-894-818B-34
; Sequence 34, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 20-MAY-1998
; APPLICATION NUMBER: US/08/894,818B
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 07-NOV-1996
; APPLICATION NUMBER: PCT/JP96/03253
; PRIOR APPLICATION DATA: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 528-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1962 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA

US-08-894-818B-34

Alignment Scores:
Pred. No.: 5.19e-31 Length: 1962
Score: 396.00 Matches: 127
Percent Similarity: 42.56% Conservative: 56
Best Local Similarity: 29.53% Mismatches: 163
Query Match: 17.58% Indels: 84
DB: 3 Gaps: 15

US-09-985-689A-7 (1-433) x US-08-894-818B-34 (1-1962)

QY 16 AsnPheGlyLeuTyGlyGlnIleValAlaValAlaAspThrGlyLeuAspThr 35
Db 454 AACITGGGATATGATGTTCTGGATCACAATAGGATAATTCACACTGGAATGAC--- 510
QY 36 GlyArgAsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAla 55
Db 511 -----GTTCTCATCCAGATCTCCAAGAAAAGTA----- 540
QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 541 ATTGGGTGGTAGATTTTGTCATGTGGAGGTATCCATACGATGACCATGACATGGA 600
QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
Db 601 ACTCATGTAGTCTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTACAAGGA 660
QY 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AepSerGlyGly 102
Db 651 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAGGTTCTAGGTGCCGATGTTCTGGAAGC 720
QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAla 122
Db 721 ATATCTACTATAATTAAGGGAGTTGAGTGGGCCGCTTGAATACAAAGATAAGTACGGAAT 780
QY 123 ArgIleHisThrAsnSerTrpGlyValProValAsnGlyAlaTyThrThrAspSerArg 142
Db 781 AAGGTCAATTAATCTTCTCTTGGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 840
QY 143 AsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 152
Db 841 CAGGCTGTAAATGAGCGGTGGGATGCTGGGATTAAGTCTGTTGCTGGCTGGAACAGT 900
QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGAACCTAACAGTATACAACTCGGTTCTCCAGCAGCTGCAAGCAAGTATTATACAGTTGA 960
QY 183 AlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAla 202
Db 961 GCC-----GTTGACCAAGTATGATGTTATAACA 987
QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProaspValMetAlaPro 222
Db 988 AGCTTCTCAAGCAGAGGGCCAACTGCAGCGCAGGCTTAAGCTGAGGTTGTTGCTCCA 1047
QY 223 GlyThrTyIleLeuSerAlaArgSerSerLeuAlaProaspSerSerPheTrpAlaAsn 242
Db 1048 GGAACCTGGAATAATGTGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 1098
QY 243 HisAspSerLysTyValAlaTyMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGCAACTCTCTCAGTAGCTGGT 1158
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
Db 1159 ATTGAGCCCTCTTGTCTCAA-----GCACACCCGAGCTGGACT 1197
QY 281 ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
Db 1198 CCAGACAAAGTAAACACAGCCCTCATAGAACTGCTCATATCGTAAAGCCAGATGAAATA 1257
QY 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312

```
Db 1258 GCCGATATAGCCTACGGTGCA-----|||||||
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329
Db 1297 AAGGCTATAAACTACGATAACTATGCAAGCTAGTGTTCACCTGGATATGTTGCCAACAAA 1356
Qy 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlySerProLeuLysIleSerLeuVal 349
Db 1357 GGCAGCAAACTACACAGTTCGGTATTAGCGGAGCTTCGGTTCGATACGCCACATATATAC 1416
Qy 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
Db 1417 TGGGACAAATGCCAAT-----AGCGACCTTGATCTT 1446
Qy 370 ValIleThrAlaProAsnGlyThrLysTy-ValGlyAsnAspPheThrAlaProTyAsp 389
Db 1447 TACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTACACCGCTACTAT--- 1500
Qy 390 AsnAsnTrpAspGlyA-GAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
Db 1501 -----GGATTGAAAAGGTTGGTTATTACACCCCACTGATGGA 1539
Qy 410 ThrTyThrValGluValGlnAlaTy-Asn 419
Db 1540 ACATGGACAATTAAGTTGTAAGCTACAGC 1569

RESULT 10
US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoskin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Alignment Scores:
Pred. No.: 5,19e-31 Length: 1962
Score: 396.00 Matches: 127
Percent Similarity: 42.56% Conservative: 56
Best Local Similarity: 29.53% Mismatches: 163
Query Match: 17.58% Indels: 84
DB: 4 Gaps: 15

US-09-985-689A-7 (1-433) x US-09-445-472-15 (1-1962)
Qy 16 AsnPheGlyLeuTyGlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThr 35
Db 454 AACTGGGATATGATGGTTCTCGAATCACAATAGGAATAGCACTGGATGAC--- 510
Qy 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAla 55
Db 511 -----GCTTCTCATCCAGATCTCCCAAGGAAAGTA----- 540
Qy 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
Db 511 -----GGATTGAAAAGGTTGGTTATTACACCCCACTGATGGA 1539
```

```
Db 541 ATTGGTGGGTAGATTTTGTCAATGGTAGGAGTTATCCATACGATGACCATGGACATGGA 600
Qy 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
Db 601 ACTCATGTAGTCTCAATAGCAGCTGGTACTGGACGACGCAAGTAATGGCAAGTACAGGGA 660
Qy 85 MetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102
Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAAGTTCTTAGGTGCGGATGTTCTTGGAGC 720
Qy 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAla 122
Db 721 ATATCTACTATATAATTAAGGAGTTGAGTGGCGCCGCTTGAATAACAAGATAAGTACGGAAT 780
Qy 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyThrThrAspSerArg 142
Db 781 AAGTTCATTAATCTTCTTGGTTCAAGCCAGAGCTCAGATGGTACTGACGCTCTAAGT 840
Qy 143 AsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlu 162
Db 841 CAGGCTGTTAATGCAAGCGTGGGATGCTGCATTAGTTGTTGGTGGCGCTGGAAACAGT 900
Qy 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
Db 901 GGACCTACACAGTATACATCGTTCTCCAGCAGCTGCAGCAAGATTATTACAGTTGGA 960
Qy 183 AlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAla 202
Db 961 GCC-----GTTGACAAGTATGATGTTATAACA 987
Qy 203 GlnPheSerSerArgGlyProThrArgAspGlyArgGileLysProAspValMetAlaPro 222
Db 988 AGCTTCTCAAGCAGAGGGCCAACTGCACAGCGGAGCTTAGCCTGAGGTGTTGCTCCA 1047
Qy 223 GlyThrTyIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1048 GGAACCTGGATAATTGCTGCCAGAGCAAGT-----GGAACCTAGCATGGGTCAACCA 1098
Qy 243 HisAspSerLysTyAlaTy-MetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1099 ATTAATGACTATTACACAGCAGCTCCTGGACATCAATGGCACTTCCTCAGTACTGGT 1158
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys----- 280
Db 1159 ATTGCAGCCCTCTTGCTCCAA-----GCACACCCGAGCTGGACT 1197
Qy 281 ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- 292
Db 1198 CCAGACAAAGTAAAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCCAGATGAATA 1257
Qy 293 AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAsp 312
Db 1258 GCCGATATAGCCTACGGTGCA-----GGTAGGGTTAATGCATAC 1296
Qy 313 LysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThrSer 329
Db 1297 AAGGCTATAAACTACGATAACTATGCCAAGCTAGTGTTCACCTGGATATGTTGCCAACAAA 1356
Qy 330 GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeuVal 349
Db 1357 GGCAGCAAACTACACAGTTCGGTATTAGCGGAGCTTCGGTTCGATACGCCACATATATAC 1416
Qy 350 TrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeu 369
Db 1417 TGGGACAAATGCCAAT-----AGCGACCTTGATCTT 1446
Qy 370 ValIleThrAlaProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyAsp 389
Db 1447 TACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTACACCGCTACTAT--- 1500
Qy 390 AsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGly 409
Db 1501 -----GGATTGAAAAGGTTGGTTATTACACCCCACTGATGGA 1539
```

QY 410 ThrThrValGluValGlnAlaTyrAsn 419
 Db 1540 ACATGACAAATTAAGTTGTAAGCTACAGC 1569

RESULT 11
 US-08-894-818B-6
 Sequence 6, Application US/08894818B
 Patent No. 6261822
 GENERAL INFORMATION:
 APPLICANT: TAKAKURA, Hikaru
 APPLICANT: MORISHITA, Mio
 APPLICANT: YAMAMOTO, Katsuhiko
 APPLICANT: MITTA, Masanori
 APPLICANT: ASADA, Kiyozo
 APPLICANT: TSUNASAWA, Susumu
 APPLICANT: KATO, Ikunoshin
 TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
 NUMBER OF SEQUENCES: 42
 CORRESPONDENCE ADDRESS:
 ADDRESS: Broadway and Neimark
 STREET: 419 Seventh Street N.W., Ste. 300
 CITY: Washington
 STATE: D.C.
 COUNTRY: United States of America
 ZIP: 20004
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/894,818B
 FILING DATE: 20-MAY-1998
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: PCT/JP96/03253
 FILING DATE: 07-NOV-1996
 PRIOR APPLICATION DATA: JP 323285/1995
 FILING DATE: 12-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Browdy, Roger L.
 REGISTRATION NUMBER: 25,618
 REFERENCE/DOCKET NUMBER: TAKAKURA-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 628-5197
 TELEFAX: (202) 737-3528
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1977 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: genomic DNA
 US-08-894-818B-6

Alignment Scores:
 Pred. No.: 2,196-30 Length: 1977
 Score: 390.00 Matches: 127
 Percent Similarity: 43.22% Conservative: 61
 Best Local Similarity: 29.20% Mismatches: 151
 Query Match: 17.32% Indels: 96
 Db: 3 Gaps: 17

US-09-985-689A-7 (1-433) x US-08-894-818B-6 (1-1977)

QY 16 AsnPheGlyLeuTyrGlyGlnGlyValAlaValAlaAspThrGlyLeuAspThr 35
 Db 454 AACTTGGGATATGATGGTCTGGATACATAGATATATGACATGGATGAC--- 510

QY 36 GlyArgAsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAla 55
 Db 1429 TCGAGCGACATCGACCTCTACCTCTACGACCCCAACGGGAACGAG---GTTGACTACTCC 1485

Db 511 -----GCTTCTCATCCAGATCTCCAAAGAAAGTA----- 540

QY 56 LeuGlyArgThrAsnAsnAlaAsn-----AspProAsnGlyHisGly 69
 Db 541 ATTGGTGGGTAGATTGTCATGTTAGGAGTTATCCATACGATGACCATGGACATGGA 600

QY 70 ThrHisValAlaGlySerValLeuGly-----AsnAlaThrAsnLysGly 84
 Db 601 ACTCATGTAGCTTCAATAGCAGCTGGTACTGGAGCAGCAAGTAATGGCAAGTCAAGGGA 660

QY 85 MetalProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGlyGlyGly 102
 Db 661 ATGGCTCCAGGAGCTAAGCTGGCGGAATTAAAGTTCTAGTTCGCGATGTTCTTGGAAAGC 720

QY 103 LeuGlyGlyLeuProAlaAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAla 122
 Db 721 ATATCTACTATAATAAGGGAGTTGAGTGGCGCTTGTATTAACAAGATAAGTACGGAATT 780

QY 123 ArgIleHisThrAsnSerTrpGlyAlaProValAsnGlyAlaTyrThrThrAspSerArg 142
 Db 781 AAGGTCATTAAATCTTCTTCTTGGTTCAAGCCAGAGCTCCGACGGAACCGACTCCCTCAGT 840

QY 143 AsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaGlyAsnGlu 152
 Db 841 CAGCGCGTCAACAACGCGCTGGAGCGCGGTATAGTCTGCGTGGCGCGCGCAACAGC 900

QY 163 GlyProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGly 182
 Db 901 GGGCGGAACACCTACACCGTGGCTACCGCGCGCGAGCAAGGTCAATACCGTCGT 960

QY 193 AlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAla 202
 Db 961 GCA-----GTTGACAGCAACGACCAATCGCC 987

QY 203 GlnPheSerSerArgGlyProThrArgAspGlyArgIleLysProAspValMetAlaPro 222
 Db 988 AGCTTCTCCAGCAGGGAGCGACCGCGAGGAGGCTCAAGCGGAAGTCTGCGCCCC 1047

QY 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
 Db 1048 GCGGTGACATCATAGCCCGCGCGCGCAGC-----GGAACCCAGCATGGGACCCCG 1098

QY 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
 Db 1099 ATAAACGACTACTACACCAAGGCGCTCTGGAACCGAGCATGCCACCGCGCAGCTTTCGGGC 1158

QY 263 ---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLys--- 280
 Db 1159 GTTGGCGGCTCATCTCCAGGCCAC-----CCGAGCTGG 1194

QY 281 ---ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 292
 Db 1195 ACCCGGACAAGGTGAAGACCGCGCTCATCGAGACCGCGCATAGTGGCCCCCAAGGAG 1254

QY 293 ---AlaAspValGlyLeuGlyPheProAsnGlyAsnGlnGlyTrpGlyArgValThrLeu 311
 Db 1255 ATAGCGGACATCGCTACGCTGG-----GTTAGGGTGAACGTC 1293

QY 312 AspLysSerLeu-----AsnValAlaPheValAsnGluThrSerProLeuSerThr 328
 Db 1294 TACAGGCGCATCAAGTACGACGACTACGCCAAGCTCACCTTACCGCTCGTGGCGGAC 1353

QY 329 SerGlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeuLysIleSerLeu 348
 Db 1354 AAGGGAAGCGCCACCCACCTTC----- 1377

QY 349 ValTrpSerAspAlaProGlySerThrThrAlaSerLeuThrLeuVal----- 364
 Db 1378 -----GAGTCAGCGGGCGGACCTTGTGACCGCCACCTTCTACTGGGACACGGGC 1428

QY 365 ---AsnAspLeuAspLeuValIleThrAlaProAsnGlyThrLysTyrValGlyAsnAsp 383
 Db 1429 TCGAGCGACATCGACCTCTACCTCTACGACCCCAACGGGAACGAG---GTTGACTACTCC 1485

Qy 384 PheThrAlaProTyrAspAsnAsnTrpAspGlyArgAsnAsnValGluAsnValPheIle 403
Db 1486 TACACCGCTACTAC-----GGCTTCGAGAAGGTGCGGTAC 1521
Qy 404 AsnAlaProGlnSerGlyThrTyrThrValGluValGlnAlaTyr 418
Db 1522 TACAACCCGACCGCGGAACCTGGACGGTCAGGTCTGTCAGTAC 1566

RESULT 12
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
US-09-000-016-3

Alignment Scores:
Pred. No.: 2,84e-24 Length: 2539
Score: 332.50 Matches: 137
Percent Similarity: 41.05% Conservative: 51
Best Local Similarity: 29.91% Mismatches: 159
Query Match: 14.76% Indels: 111
DB: 3 Gaps: 23

US-09-985-689A-7 (1-433) x US-09-000-016-3 (1-2539)
Qy 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37

Db 998 GGCTACACGGCAGGCGGTGAAGATCCCGTCTCGACACCGGTCTCGACACGAGC--- 1054
Qy 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCCGGACCTGAGGGCGGGGTGACCGCGTCAAGAACTTCACC 1099
Qy 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCCGCGCGCGCGCGGCGGACAGGTGGCGACCGGACCCACCTCGCTCGATCGCGGCG 1159
Qy 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCAACGGCGCGCCAGTCCAAAGGCGCAAGTACAAAGGGCGCTCGACCCGCGCGGATCCTC 1219
Qy 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110
Db 1220 AACGGCAAGGTCCTCGACACTCCGGTTTGGGACACACTCCGGCATCTCCGCGGATG 1279
Qy 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1280 GAGTGGCGCGCGCGCAG-----GGCGCGACGTCGTACCATGAGCTGGGC 1327
Qy 131 Ala-----ProValAsnGlyAlaTyrThrThrAspSerArgAsn 143
Db 1328 GGCATGGACACACCGGAGACCGACCCGCTGGAGCGGCG----- 1366
Qy 144 ValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 CTCGACAAGCTGTCGCGCGAGAGGGGCTCTGTCCTCCATCGCGCGCGGCAACGAGGCG 1426
Qy 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTGC-----ATCGGTTCCGCGGACGCGGACGCGCGCTCCACCTCGCGCGGCC 1480
Qy 184 ThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 203
Db 1481 -----GTCGACGACAGGACAGCATCGCGAC 1507
Qy 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMerAlaPro 222
Db 1508 TTCTCTCTCACCGCGCCCGCTCGCGACGCGGCCCATCAAGCGCGGAGCTCACCGCTCCC 1567
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCACGCGCGCTCGCGGAGGCGGACGACATCGCGCGGAGGTCGTGAG 1627
Qy 243 HisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACCGCGCGGTACATGACCATCTCCGGCACGTCGATGGGACCCCGCACGTCGCGGCG 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282
Db 1688 GCGCGCGCGCTCTCTGAAGCAGCAG-----CACCCCGAC 1720
Qy 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db 1721 TGGACCTCGCGCGAACTGAAGGCGCGCTCACCGGCTCCACCAAGGCGGCAAGTACACC 1780
Qy 300 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db 1781 CCG---TTCGACGAGGGTTCCGGCGCGATCCAGGCGGCAAGGCGCTCCAGCAGACCGGTG 1837
Qy 320 ValAsnGluThrSerProLeuSerThrSer----- 329
Db 1838 ATC-----GCCGACCGCGTCTCGGTGAGCTTCGGGGTCCAGCAGTGGCGGCACACCGAC 1891
Qy 330 -----GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeu 344
Db 1892 GACGAGCGGTCAACCAAGCAGCTGACCTACCGCAACCTCGGCACCCACGACGACGTCGCTG 1951
Qy 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354

QY 320 ValAsnGluThrSerProLeuSerThrSer----- 329
Db 1838 ATC-----GCCACCGGCTCTCGTGAGCTTCGGGCTCCAGAGTGGCGGCACACCGAC 1891
QY 330 -----GlnLysAlaThrTySerPheThrAlaGlnAlaGlyLysProLeu 344
Db 1892 GACGAGCGGTCACCAAGCAGCTGACCTACCGCAAGCTCGGCACCCAGGACGTCAGCTG 1951
QY 345 LysLeuSerLeuValTrpSerAsp-----AlaPro----- 354
Db 1952 AGCTGACGTCGACCGCCACCGACCCCAAGGCAAGGCGGCGGCGGCTTCTTCAGC 2011
QY 355 ---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuValLeuThrAla 373
Db 2012 CTGGCGCCACCG-----GTGACCGTC 2035
QY 374 ProAsnGlyThrLysTyValGlyAsnAspPheThrAlaProTyTrpAspAsnTrpAsp 393
Db 2036 CCGCGCGCGGCGAGC---GCCTCGTCGATACACCGC-----GACACCGGCTCGGC 2086
QY 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyThr 412
Db 2087 GGCACGGTGACGCGGCTACTCGGCTACGTGTGCGC---ACGGCGCGCGGCGAGCG 2143
QY 413 Val-----GluValGlnAlaTyTrpAsnValProVal 422
Db 2144 GTCGCGACGCGCGCGCGGTGCAGCGGAGTGCAGTCGACGAGTACCGTC 2197

RESULT 14
US-09-000-016-1
; Sequence 1, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV
; TITLE OF INVENTION: ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:

ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 1.09e-23 Length: 2809
Score: 327.50 Matches: 136
Percent Similarity: 40.83% Conservative: 51
Best Local Similarity: 29.69% Mismatches: 160
Query Match: 14.54% Indels: 111
DB: 3 Gaps: 23
US-09-985-689a-7 (1-433) x US-09-000-016-1 (1-2809)
QY 18 GlyLeuTyGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGTACGACGCGGAGGGGTGAGATCGCTCTGGACACCGGTGTGCACGAGC--- 1054
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyThrAlaLeuGly 57
Db 1055 -----CATCCGACCTGAAGGCGGGTGCACCGGCTCCAAAGAACTTACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCGCGCGCGCGCGCGGACAGGTGGCGCACCGCACCCACGCTCGCTCGATCGCGCG 1159
QY 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGCGCGCGCGCGTCCAAAGGGCAAGTACAAAGGCGTGCACCGCGCGCGATCTC 1219
QY 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110
Db 1220 AACGCAAGGTCTCTGACGACTCCCGTTTCGGCGACGACTCGGCGATCTCTCGCGCGATG 1279
QY 111 GlnThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1280 GAGTGGCGCGCGCGAG-----GGCGCGAGCTGTGTCAATAGCTGAGCTGGGC 1327
QY 131 Ala-----ProValAsnGlyAlaTyThrThrAspSerArgAsn 143
Db 1328 GGCATGGACACACCGGAGACCGCGCTGGAGCGCGG----- 1366
QY 144 ValAspAspTyTrpValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGly 163
Db 1367 GTCGACAAGCTGTCCCGCGAGAGGGCGTCTTCGCGCATCGCGCGCGCAACGAGGCG 1426
QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db 1427 CCGGAGTCG-----ATCGTTTCGCGCGCGAGCGCGCGCTTCCCGTGGCGCGC 1480
QY 184 ThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAlaGln 203
Db 1481 -----GTGACGACAGGACAAAGTCCGCGAC 1507
QY 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMetAlaPro 222
Db 1508 TTCTCTCCACCGCGCGCGCGCTCGCGCGAGCGCGCATCAAGCGCGAGCTCACCGCTCC 1567
QY 223 GlyThrTyThrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGGACATCAGCGCGCGCTTCGCGGAGGGGCAACGATCGCGCGAGGTCGCGTGA 1627

QY 243 HisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACGGCGCGCTACATGACCATCTCCGGCAGCTGATGCGCAGCCCGCAGCTCGCGGGC 1687
QY 263 AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyValThrProLysProSer 282
Db 1688 GCGGCGCGCTCTGAGCAGCAG-----CACCCTCGAC 1720
QY 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db 1721 TGGACCTCCGGCGGACTGAAGCGCGGCTCACCCTCCACAGGCGCGCAAGTACACC 1780
QY 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db 1781 CCG---TTCGAGCAGGTTCCGGCGGATCCAGGCCGACAGGCGCTCCAGCAGACCGTG 1937
QY 320 ValAsnGluThrSerProLeuSerThrSer----- 329
Db 1838 ATC-----GCCGACCCCGTCTCGGTGAGCTTCGCGTCCAGCAGTGGCGGCACCGCAC 1891
QY 330 -----GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeu 344
Db 1892 GACGAGCGCGTACCAAGCAGCTGACCTACCGCACTCGGCACCCAGGACGTACGCTG 1951
QY 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354
Db 1952 AAGCTGAGCTGACCGCCACCGCAGCCAGGCGGCAAGGCGCGCGCGGCTTCTTCACG 2011
QY 355 ---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 2012 CTGGCGCGCACACG-----GTGACCGTC 2035
QY 374 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 393
Db 2036 CCGCGCGCGCGCAGC---GCTCCGTCGACATGACCGCC-----GACACCGCGCTCGGC 2086
QY 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412
Db 2087 GGCACGGTGGAGCGCGGTACTCGCGTACGTGTCGCC---ACGGCGCGCGCGCAGACG 2143
QY 413 Val-----GluValGlnAlaTyrAsnValProVal 422
Db 2144 GTCCGCGCGCGCGCGGTGCGAGCGGAGTGTGAGTGTGACGAGTGTGACGAGTGTGACGCGTC 2197

RESULT 15

US-09-514-340-1
Sequence 1, Application US/09514340
Patent No. 6361987
GENERAL INFORMATION:
APPLICANT: Akira ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROXYRINDINE DE
ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/514,340
FILING DATE: 28-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: 09/000,016
FILING DATE: January 30, 1998

ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN: <Unknown>
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1

Alignment Scores:
Pred. No.: 1,09e-23 Length: 2809
Score: 327.50 Matches: 136
Percent Similarity: 40.83% Conservative: 51
Best Local Similarity: 29.69% Mismatches: 160
Query Match: 14.54% Indels: 111
DB: 4 Gaps: 23

US-09-985-689A-7 (1-433) x US-09-514-340-1 (1-2809)

QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 998 GGCTACGACGCGCAAGGCGGTGAAGATCGCGTCTCGACACCGGTCTCGACACGAGC--- 1054
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 1055 -----CATCGGACCTCAAGGCGCGGTGACCGCGTCCAGACATTCACC 1099
QY 58 ArgThrAsnAsnAlaAsnAspProAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 1100 GCCGCGCGCGCGCGCGCGCAAGGTTGGCCCGCACCGCACCGCTCGATCGCGCG 1159
QY 78 GlyAsnAlaThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 92
Db 1160 GGCACGGCGCGCCAGTCCAGGGCAAGTACAGGGGTGCGCACCCCGCGCGGATCTC 1219
QY 93 PheGlnSerIleMet---AspSerGly---GlyGlyLeuGlyGlyLeuProAlaAsnLeu 110
Db 1220 AACGCGCAAGGTCTCTCGACGACTCCGGTTTCGGGCGACGACTCCCGCATCTCGCGCGCATG 1279
QY 111 GlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGly 130
Db 1280 GAGTGGGCGCGCGCAG-----GGCGCGGACGTGCTCAACATGACCTGGGCT 1327
QY 131 Ala-----ProValAsnGlyAlaTyrThrThrAspSerArgAsn 143
Db 1328 GGCATGGACACACCGGAGACCGCCGCTGGAGCGCGCG----- 1366
QY 144 ValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGly 163
Db 1367 GTCGACAGGTGTCGCGCGGAGAGGCGCTCTCTGCCATCGCGCGCGCACGAGGCG 1426
QY 164 ProGlySerGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAla 183
Db ----- 183

Db 1427 CCGGAGTCG-----ATCGGTTCGCGCGGAGCGGCGCGCCCTCACCCTCGCGCGC 1480
Qy 184 ThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGln 203
Db 1481 -----GTGACGACCAAGGACCAAGCTCGCCGAC 1507
Qy 204 PheSerSerArgGlyPro---ThrArgAspGlyArgIleLysProAspValMetAlaPro 222
Db 1508 TTCTCTCCACCGCGCCCGCTCGGCGGACGCGGCCATCAAGCCGCGGTCCACCGCTCC 1567
Qy 223 GlyThrTyrIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsn 242
Db 1568 GCGGTGACATCAGCGCGCTCGCGGAGGCAACGACATCGCGCCGAGGTCGGTGAG 1627
Qy 243 HisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly 262
Db 1628 GGACCGCGCGGTACATGACCATCTCCGGCACGTGATGCGACCCCGCACGTCGCGGGC 1687
Qy 263 AsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyValThrProLysProSer 282
Db 1688 CCGCGCGCCCTCTGAAGCAGCAG-----CACCCCGAC 1720
Qy 283 LeuLeuLysAlaAlaLeu-----IleAlaGlyAlaAlaAspValGlyLeuGlyPhe 299
Db 1721 TGGACCTCCGCGGAGCTGAAGGCGCGCTCACCGCTCCACCAAGGCGGCAAGTACACC 1780
Qy 300 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaPhe 319
Db 1781 CCG---TTCGAGCAGGGTTCGGCGCGGATCCAGGCGCAAGGCGCTCCAGCAGACCGTG 1837
Qy 320 ValAsnGluThrSerProLeuSerThrSer----- 329
Db 1838 ATC-----GCCGACCCGCTCTCGGTGAGCTTCGGCGTCCAGCAGTGGCGCGCACCGAC 1891
Qy 330 -----GlnLysAlaThrTyrSerPheThrAlaGlnAlaGlyLysProLeu 344
Db 1892 GACGAGCGGTCCACCAAGCAGCTGACCTACCGCAACCTCGGCACCCAGGACGTCACGCTG 1951
Qy 345 LysIleSerLeuValTrpSerAsp-----AlaPro----- 354
Db 1952 AAGCTGACGTGACCGCGCACCGCCCAAGGCGAAGCGCGCCCGCGGCTTCTTCACG 2011
Qy 355 ---GlySerThrThrAlaSerLeuThrLeuValAsnAspLeuAspLeuValIleThrAla 373
Db 2012 CTGGGCGCACCCAG-----GTGACCGTC 2035
Qy 374 ProAsnGlyThrLysTyrValGlyAsnAspPheThrAlaProTyrAspAsnAsnTrpAsp 393
Db 2036 CCGCGGCGCGCAGC---GCCTCCGTGACATGACCGCC-----GACCCCGGCTCGGC 2086
Qy 394 GlyArg---AsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThr 412
Db 2087 GGCACGGTGGACGGCGGTACTCGCGGTACGTGTCGCC---ACGGCGGCGGCGCAGACG 2143
Qy 413 Val-----GluValGlnAlaTyrAsnValProVal 422
Db 2144 GTCCGACGCGCGCGGTGCGCGGCGGAGTCTGAGTCTGAGTCTGAGTCTGAGTCTGAGTCT 2197

Search completed: April 4, 2004, 12:04:03
Job time : 92.4512 secs